Small-molecule probes to explore cancer

Small-molecule probes to explore cancer

A dissertation presented

by

Giannina Ines Schaefer

to

The Department of Chemistry and Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemistry

Harvard University

Cambridge, Massachusetts

April 2014

Dissertation Advisor: Giannina Ines Schaefer

Professor Stuart L. Schreiber

Small-molecule probes to explore cancer

Abstract

Small molecules play important roles in therapeutics and drug discovery. Significant progress has been made by the chemical biology community to discover small-molecule probes to explore biological processes and to treat disease. This thesis describes both the discovery of novel probes for the Hedgehog (Hh) pathway and the application of small molecules in identifying cancer dependencies.

In a phenotypic screen for inhibitors of Hh signaling, a pathway sometimes deregulated in cancer, several previously annotated molecules were identified that, in addition to their original targets, were shown to have activity in this pathway. Furthermore, two potent small-molecule inhibitors, BRD50837 and BRD9526, were discovered. Analysis of structure–activity relationships (SAR) demonstrated striking stereochemistry-based SAR. This feature suggested a specific and selective interaction of these compounds with their cellular target(s). Further studies revealed that their mechanism of action displayed similarities to that of cyclopamine, a commonly used Hh probe that targets the Smoothened receptor, and yet differed strikingly in other aspects. Early insights into the probes' mechanisms of action shed light onto the nature of these novel compounds.

To advance the application of small molecules in cancer, an unbiased screen was performed using 242 genomically characterized cancer cell lines (CCLs) that were profiled using a set of 354 small molecules to discover novel oncogene and non-oncogene dependencies. Enrichment correlations between small-molecule sensitivity and genetic features were calculated. This allowed for known dependencies to be confirmed and generated several novel hypotheses. A public resource was created based on these efforts (http://www.broadinstitute.org/ctrp).

To address the challenge of interpreting genomic alteration/compound sensitivity relationships, genomic characterizations were prioritized using recurrent mutations, overlap with patient data, and annotation of fusion genes. This prioritized approach demonstrated that small-molecule sensitivity can be differential across mutations in the same gene, and allowed for identification of novel and specific dependencies. For example, pan- and PI3K-α-specific inhibitors appear to be more potent in *PIK3CA* E542K-mutated CCLs than other *PIK3CA*-mutated CCLs.

Table of Contents

For my parents.

MY head knocks against the stars.

My feet are on the hilltops.

My finger-tips are in the valleys and shores of universal life.

Down in the sounding foam of primal things I reach my hands and play with pebbles of destiny.

I have been to hell and back many times.

I know all about heaven, for I have talked with God.

I dabble in the blood and guts of the terrible.

I know the passionate seizure of beauty

And the marvelous rebellion of man at all signs reading "Keep Off."

My name is Truth and I am the most elusive captive in the universe.

- Carl Sandburg ("*Who am I?*")

Acknowledgements

A doctoral thesis is never just the work of one person. While one person puts all the pieces together, the final product is a concoction of many people's thoughts, comments, encouragement, and support. I would like to first and foremost thank my advisor, Stuart Schreiber, for all his guidance throughout the years. Your enthusiasm for science has greatly motivated me and I was always happy that I made the decision to join Harvard and your group. I consider myself lucky to be able to learn from you both, scientific and personal skills.

My gratitude extends to everyone who I had the pleasure of working with and being friends with in the lab, at Harvard, and at the Broad. I am very grateful to my collaborators, coworkers and those who have kindly shared materials with me. I had the chance to interact with a very special group of people and enjoyed each conversation tremendously. Thank you to everyone on the CTD^2 project team and those who have worked with me on the Hedgehog project – it was great to witness both projects grow throughout the years.

Thank you in particular to Leslie Aldrich, Jaime Cheah, Jeremy Duvall, Sigrun Gustafsdottir, Angela Koehler, Max Majireck, Pat Mark, Frances Neville, Josh Paulk, Jose Perez, Sally Ricupero, Aly Shamji, Michelle Stewart, and Bridget Wagner. Leslie and Max, our conversations about synthetic chemistry have greatly helped me – as hood neighbors and throughout 3175. Jaime, you have taught me countless biology techniques and the daily Starbucks runs have been a lot of fun. Aly, Jose, and Jeremy, your scientific input has greatly helped me decide how to pursue my projects. Jose,

thank you especially for your support in Hedgehog-related questions through all these years. Sigrun and Michelle, you were always there as my desk mates, the first ones to answer questions, no matter how trivial and to give me advice I needed. Michelle, it has been a great time to work with you together on the recurrent mutation project. Angela, Michelle, and Bridget, your advice throughout the years of how to present my findings has been priceless. Thank you especially for taking the time to read through the drafts of this thesis. Pat, Frances, and Sally, this lab would not work as well without you and you always take the time to also listen to all the personal issues as well. Josh, I would not have even started my Hedgehog project had it not been for your encouragement – one of the best decisions I made.

I would like to thank my professors in college, especially Prof. Werner Nau and Prof. Thomas Nugent. Without your teachings and guidance I would not even have had the chance to start this journey and your continued support has greatly helped me. Thank you also to Prof. Dan Kahne and Prof. Alan Saghatelian, my committee members. Your kind advice and suggestions have always helped me see the big picture and move towards this thesis.

None of this would have been possible without the emotional support from my family and friends. I am deeply thankful to my family, especially my parents Petra and Dietmar Schaefer and my sisters Darja and Raika Schaefer – you have always challenged me to think big, supported me in my endeavors and never once doubted it could be done. Helga Hensing, you were also always there through all these years and I have learned many important things from you. To all of my friends near and far: you have always supported me. Robert Day, Anca Dragan, Ruixuan Gao, Steffi Hamann,

Stefanie Janzen, Björn Kaschubick, Clare Kim, Shelarese Ruffin, Esther Singer, Benjamin Stevenson, Robin Sussman – thank you in particular for the hours of phone calls and endless talks over beer and coffee which helped tackle the ups and downs of graduate school. Finally, thank you to my Taekwondo family, especially Master Park, Master Jang, Master Lee, and Yoshi Fujita. You have taught me extremely valuable lessons that were also applicable off the mat and have helped me with the graduate school challenge.

List of abbreviations

Amino acid abbreviations

A	Ala	alanine	M	Met	methionine
C	Cys	cysteine	N	Asn	asparagine
D	Asp	aspartate	P	Pro	proline
E	Glu	glutamate	Q	Gln	glutamine
F	Phe	phenylalanine	R	Arg	arginine
G	Gly	glycine	S	Ser	serine
H	His	histidine	T	Thr	threonine
I	Ile	isoleucine	V	Val	valine
K	Lys	lysine	W	Trp	tryptophan
L	Leu	leucine	Y	Tyr	tyrosine

Gene and protein names

ACVR2A	-	activin receptor type-2A
AKAP12	-	A kinase (PRKA) anchor protein 12
AKAP9	-	A kinase (PRKA) anchor protein 9
AKT	-	RAC serine/threonine-protein kinase
ALK	-	ALK tyrosine kinase receptor
ALPK2	-	alpha-protein kinase 2
BCR/ABL	-	breakpoint cluster region/ Abelson murine leukemia viral oncogene homolog 1 fusion gene
BRAF	-	serine/threonine-protein kinase B-raf
BRCA1	-	breast cancer 1, early onset
BRCA2	-	breast cancer 2, early onset
CDC25C	-	M-phase inducer phosphatase 3
CDK4	-	cyclin-dependent kinase 4
CDKN2A	-	cyclin-dependent kinase inhibitor 2A
CHD1	-	chromodomain-helicase-DNA-binding protein 1
CLTCL1	-	clathrin heavy chain 2
CRAF	-	RAF proto-oncogene serine/threonine-protein kinase
CREB3L2	-	cyclic AMP-responsive element-binding protein 3-like protein 2
CTNNB1	-	catenin beta-1
CXCR2	-	C-X-C chemokine receptor type 2

EGFR	-	epidermal growth factor receptor
ERBB2	-	receptor tyrosine-protein kinase erbB-2
EWSR1	-	RNA-binding protein EWS
FLI1	-	friend leukemia integration 1 transcription factor
Gli	-	GLI family zinc finger
GPR112	-	probable G-protein coupled receptor 112
GPR39	-	G-protein coupled receptor 39
GSTM	-	glutathione S-transferase mu
HGF	-	hepatocyte growth factor
Hh	-	hedgehog
HRAS	-	GTPase HRas
hTERT	-	telomerase reverse transcriptase
ITPR2	-	inositol 1,4,5-trisphosphate receptor type 2
KAT6B	-	histone acetyltransferase KAT6B
KRAS	-	GTPase KRas
MAML3	-	mastermind-like protein 3
MAP3K1	-	mitogen-activated protein kinase kinase kinase 1
MAP3K14	-	mitogen-activated protein kinase kinase kinase 14
MEK	-	dual specificity mitogen-activated protein kinase kinase
MITF	-	microphthalmia-associated transcription factor
MLLT3	-	protein AF-9
MTOR	-	serine/threonine-protein kinase mTOR
MYC	-	Myc proto-oncogene protein
NAMPT	-	nicotinamide phosphoribosyltransferase
NCOA3	-	nuclear receptor coactivator 3
NEK3	-	serine/threonine-protein kinase Nek3
NOTCH1	-	neurogenic locus notch homolog protein 1
NPM1	-	nucleophosmin
NR1H2	-	oxysterols receptor LXR-beta
NRAS	-	GTPase NRas
PARP	-	poly ADP ribose polymerase
PI3K	-	phosphoinositide 3-kinase (protein)
PIK3C2G	-	phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit gamma
PIK3CA	-	phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform
POLA1	-	DNA polymerase alpha catalytic subunit
POLA2	-	DNA polymerase alpha subunit B
POLE	-	DNA polymerase epsilon catalytic subunit A
PRKDC	-	DNA-dependent protein kinase catalytic subunit
Ptch	-	Patched

RECQL4	-	ATP-dependent DNA helicase Q4
Rho	-	Ras homolog gene family
RPL22	-	60S ribosomal protein L22
Shh	-	Sonic Hedgehog
Smo	-	Smoothened
STK11	-	serine/threonine-protein kinase STK11
SuFu	-	Suppressor of Fused
SV40	-	simian vacuolating virus 40
TP53	-	cellular tumor antigen p53
VEGFC	-	vascular endothelial growth factor C
WNT	-	wingless-related integration site

General abbreviations

ADP	-	adenosine diphosphate
ATP	-	adenosine triphosphate
AUC	-	area under curve
BCC	-	basal cell carcinoma
BCS	-	bovine calf serum
BH3.DMS	-	borane dimethyl sulfide complex
Boc	-	tert-butoxycarbonyl
CCL	-	cancer cell-line
CCLE	-	Cancer Cell Line Encyclopedia
CGA	-	Cancer Genome Analysis
CM	-	Shh-conditioned medium
CML	-	chronic myelogenous leukemia
CNV-H	-	high-copy number
COSMIC	-	Sanger Catalogue of Somatic Mutations in Cancer
CRISPR	-	clustered regularly interspaced short palindromic repeats
CTRP	-	Cancer Therapeutics Response Portal
DCM	-	dichloromethane
DDQ	-	2,3-dichloro-5,6-dicyanobenzo-quinone
DIPEA	-	N,N-diisopropylethylamine
DMEM	-	Dulbecco's Modified Eagle Medium
DMF	-	dimethylformamide
DNA	-	deoxyribonucleic acid
DOS	-	diversity-oriented synthesis
FDA	-	Food and Drug Administration
GANT-61	-	GLI Antagonist 61

GCG	-	(-)-gallocatechin-3-monogallate
GTPase	-	guanine triphosphatase
HDAC	-	histone deacetylase
HF	-	hydrogen fluoride
IE	-	BRD50837's inactive enantiomer
iTRAQ	-	isobaric tags for relative and absolute quantification
MLSCN	-	Molecular Libraries Screening Centers Network
MS	-	mass spectrometry
MUT	-	any mutation call
Onco	-	Oncomap mutant calls
PBS	-	phosphate buffered saline
PMB	-	paramethoxybenzyl ether
PyBOP	-	(benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate
RNA	-	ribonucleic acid
RNAi	-	RNA interference
SAG	-	Smoothened Agonist
SAR	-	structure–activity relationships
STAB	-	sodium triacetoxyborohydride
TBS	-	*tert*-butyldimethylsilyl
TBSOTf	-	*tert*-butyldimethylsilyl trifluoromethanesulfonate
TBST	-	tris-buffered saline with tween 20
TCGA	-	The Cancer Genome Atlas
TES	-	all targeted-exome sequencing mutant calls
TES-A	-	targeted-exome sequencing, non-neutral missense mutations
TFA	-	trifluoroacetic acid
THF	-	tetrahydrofuran

Chapter One

Introduction to small-molecule probe development

Small-molecule probe development

Small molecules are omnipresent in nature. They play important but diverse roles in biology including, but not limited to, functioning as signaling molecules,[1,2] metabolites,[3] and even warfare agents between different biological species.[4,5] They are therefore a great part in the chemistry of life and have been described as an important piece of the central dogma of life.[6] It is thus unsurprising that they have been used both as tools in biology and as drugs in medicine.[7] As probes, small molecules have helped to elucidate the underlying biology of cells and of diseases: for example, colchicine and taxol helped illuminate the cytoskeleton,[8,9] and cyclosporine A, FK506, and rapamycin helped understand immune system signaling and how to manipulate it.[10] Some probes even subsequently become FDA-approved drugs, such as in the case of the immunosuppressant rapamycin.[11] Small-molecule probes, therefore, play an important role in our understanding of biology and in drug discovery, and their development will continue to help us treat patients.

Small-molecule probe development has become an important part of medicinal research and is done by both academia and pharmaceutical industry. Because these tools are being used to generate and prove hypotheses about biology and diseases, it is important that they are reliable in their mechanism of action. Therefore, quality-control measures have been defined. For example, the Molecular Libraries Screening Centers Network (MLSCN) defined the criteria for a small molecule to be a probe to be: it must have a potency of at least 100 nM, selectivity of at least tenfold over related targets, and aqueous solubility. It furthermore needed to be an improvement over already existing probes.[12] Other important criteria are cell permeability and potency in cellular assays to

ensure target engagement in biological systems as well as *in vitro*.[13] Drugs do not necessarily fulfill the stringent criteria of selectivity, a characteristic that might even be beneficial if several small-molecule/target interactions are needed to create the desired effect in the human body.[12] However, in order to dissect cellular biology in the most detail possible, probes are more useful the more they fulfill the criteria.

The discovery of small-molecule probes, like that of drug candidates, can be done using high-throughput screening. Screens can, very broadly, be divided into two classes: biochemical and cell-based screens.[14] Biochemical screens generally are of a more directed nature: the protein to be targeted will be screened for directly, for example by using purified enzymes or small-molecule microarrays.[15] Advantages for this type of screen are that the target is known beforehand and therefore, mechanism of action is known. Disadvantages include that the compounds may not be cell-permeable or active in the cell-based system, or that the compounds have different or additional effects and targets in a cellular setting, thus compromising their usefulness as tools. Cell-based screens allow for modulators of different proteins in a pathway to be found (using reporter-gene-based assays for example), or even for modulators of a phenotype (phenotypic assays), which could in theory result in modulators of any member of the proteome. These approaches allow for more targets within a pathway to be screened simultaneously and they also allow for the discovery of previously not-considered targets. Their major disadvantage lies in the requirement for subsequent target identification which can be difficult and time-consuming. While there is no clear prescribed method to identify the target, advances in proteomics, genetics, and bioinformatics, have led to a toolbox to manage this task.[16,17] While these methods are

important to elucidate the mechanism-of-action of a probe or drug, they can also help to determine possible off-targets.

The Sonic Hedgehog signaling pathway

One example where probe development and pathway elucidation have gone hand-in-hand is in the case of the Sonic Hedgehog (Shh) signaling pathway. This signaling pathway is necessary for embryonic development and plays an important role in growth, patterning, and morphology of insects and vertebrates.[18,19] Deficiencies in Hedgehog (Hh) signaling during development can cause diseases such as holoprosencephaly.[20] Additionally, aberrant activation of the pathway through somatic mutations in genes encoding members of the pathway are known to drive the development and maintenance of several cancers, such as basal cell carcinoma (BCC), pancreatic cancer, and medulloblastoma.[21-24]

Research on the pathway has been ongoing since the 1970s when researchers identified genes which, when mutated, affected the segmental pattern of *Drosophila melanogaster* larvae.[25] Since then much research into its members and signaling mechanism has been done, but understanding of the pathway is still not complete and the picture is becoming increasingly complex. In the classical “canonical” signaling model the pathway is activated by one of three extracellular secreted proteins: Desert Hedgehog, Indian Hedgehog, or, most commonly, Sonic Hedgehog (Shh) (Figure 1.1a). This protein binds the transmembrane receptor Patched (Ptch). In the absence of ligand binding, Ptch represses Smoothened (Smo), a G-protein coupled transmembrane receptor (Figure 1.1b). In a still unknown way, the Shh/Ptch complex leads to a

derepression of Smo, resulting in its translocation to the primary cilium. The transcriptional regulator of the pathway, Gli, is usually sequestered in a repressor complex with, among others, Suppressor of Fused (SuFu), which results in degradation to its inactive form. Translocation of Smo to the primary cilium results in a release of Gli from the repressor complex and translocation of the activated transcription factor to the nucleus where genes involved in cell proliferation and differentiation are activated.[21,26,27]

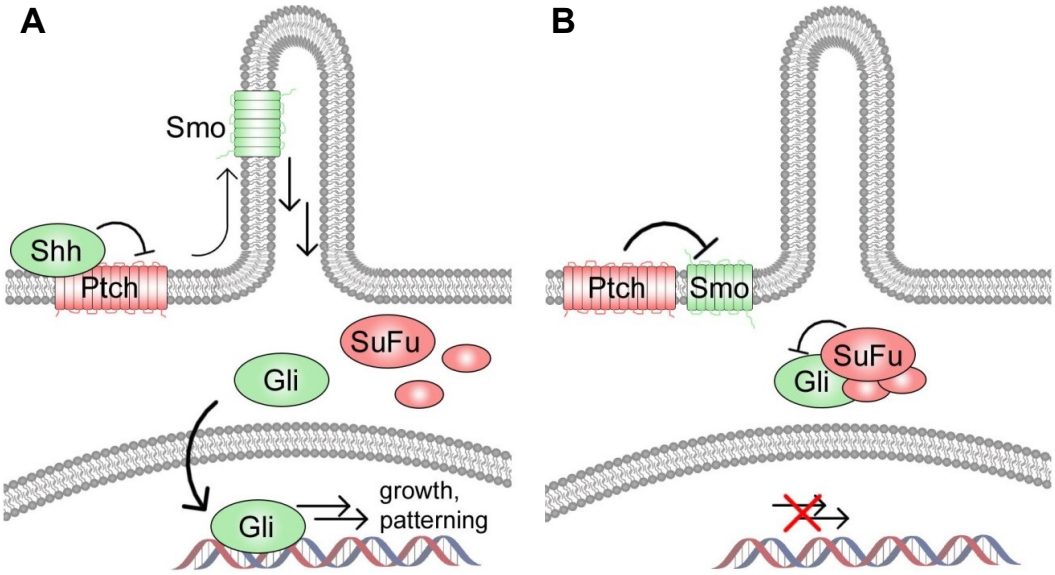
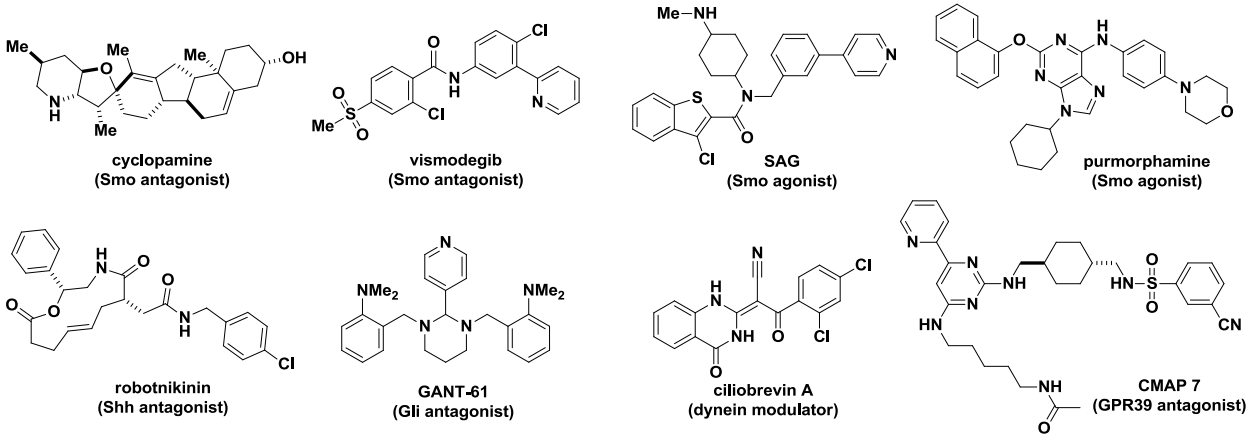


Figure 1.1. Canonical signaling of the Hedgehog pathway in the a) active and b) inactive state.

The first small-molecule inhibitor of the pathway to be discovered was cyclopamine, a natural product found in *Veratrum Californicum* and a Smo inhibitor (Scheme 1).[28] More Smo inhibitors were subsequently identified, the most notable of which is vismodegib, an FDA-approved treatment for BCC.[29,30] Additionally, small-molecule activators of the pathway have been found, such as SAG and purmorphamine which also act on Smo.[31,32] As an example of a targeted screen, a small-molecule microarray setup has led to an antagonist of Shh, robotnikinin.[33] In cell-based assays,

on the other hand, inhibitors of the Gli/DNA interaction have been found (e.g. GANT-61) and small molecules which later proved to inhibit the pathway by modulating the motor protein dynein (ciliobrevin A).[34,35] Recently, a small molecule inhibiting Hh signaling was discovered which was found to target GPR39, a G protein-coupled receptor with previously no known connection to the Hh pathway.[36] Elucidation of these compounds' mechanism of action therefore has furthered the understanding of the pathway's signaling steps.

Scheme 1.1. Structures of Hedgehog pathway modulators.



In addition to the 'canonical' Hh signaling, more recent evidence has shown that the pathway also promotes 'noncanonical' signaling.[37] Examples of this signaling include Gli-independent activation of Rho small GTPases through Smo and pro-apoptotic signaling through Shh and Ptch, independent of Smo.[38-40] Further complexity is added by the observation that small-molecule modulators of the pathway can act on the same protein and have different cellular outcomes. For example, cyclopamine promotes Smo accumulation in the primary cilium while this is prevented by vismodegib.[41,42] Even more strikingly, purmorphamine and cyclopamine bind in the

same Smo binding pocket, but purmorphamine activates the pathway while cyclopamine inhibits it.[28,32]

Cancer dependencies

The development of vismodegib as a treatment for BCC was based on the fact that nearly 100% of patients with sporadic BCC have mutations causing faulty activation of the Hh pathway[43-45] and the premise that cancer cells become "dependent" on their activated growth signals. Cancers are the result of a series of somatic mutations of oncogenes, leading to uncontrolled growth. Some of these mutations are required to continue growing; removing the signal from these mutations causes the cancer cells to arrest their growth or die. As healthy cells generally do not depend on signals such as the Hh signal, they would not be as strongly affected by removal or inhibition of this oncogene (Figure 1.2a).[46] This concept, known as oncogene dependency, is a relatively recent concept and has been used to create targeted chemotherapies with limited toxicity profiles. Examples of therapies targeting such oncogene dependencies that exist in the clinic are vemurafenib, an inhibitor of V600E-mutated *BRAF* approved for metastatic melanoma, or imatinib, an inhibitor of the BCR/ABL fusion protein and c-kit kinases approved for chronic myelogenous leukemia and gastrointestinal tumors.[47,48] In addition to oncogene dependencies, it appears that there are also non-oncogene dependencies where cancer cells have become dependent on certain housekeeping or signaling genes as a result of their genetic mutations (Figure 1.2b). Examples of this type of dependency are breast cancer cell lines with mutations in *BRCA1* or *BRCA2*

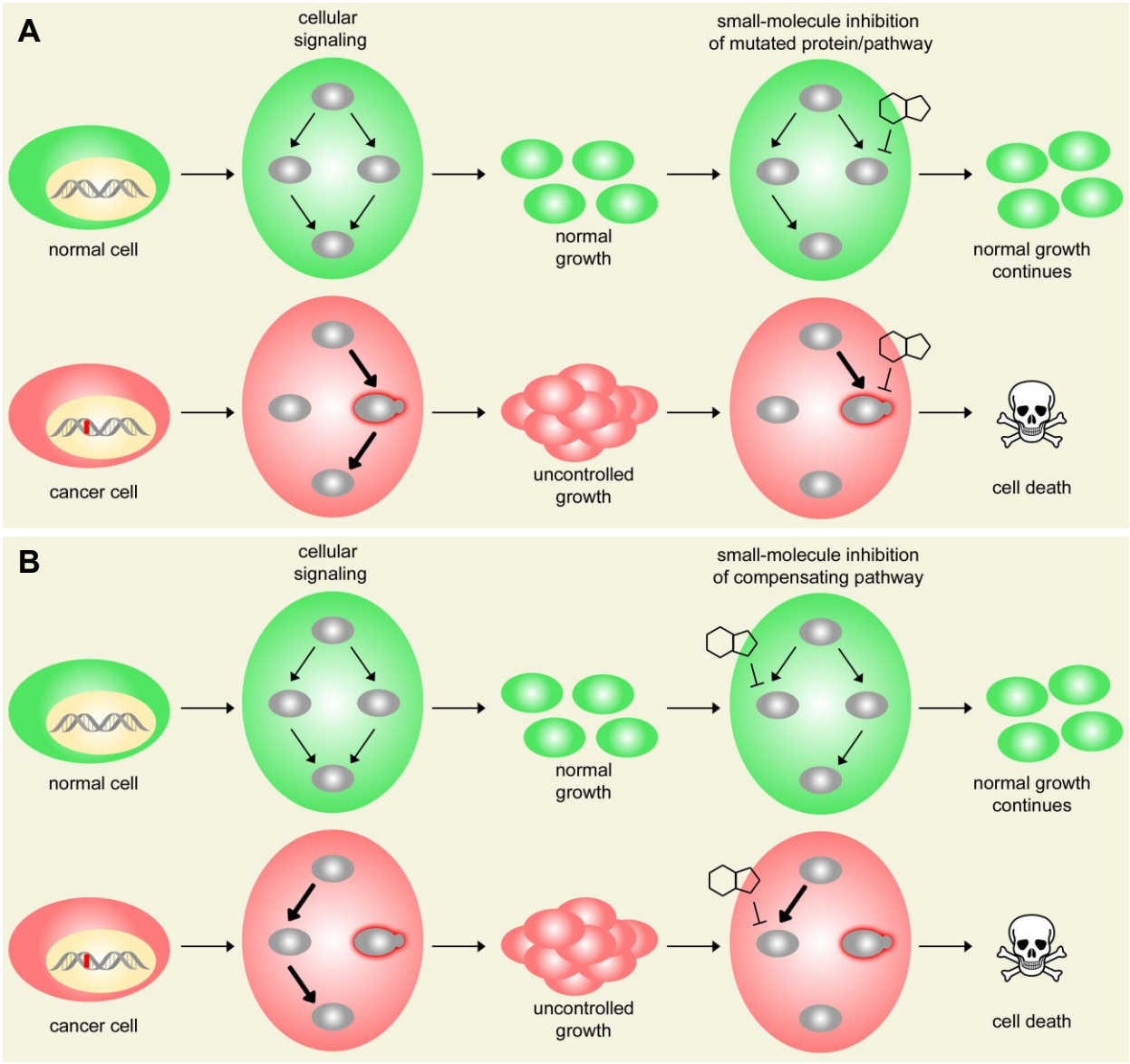


Figure 1.2. Mutations in cancer cells cause these cells to rewire their signaling around the overactivated or damaged protein. As a result of this, cancer cells can become dependent on the mutated protein/pathway (a), or another, compensating pathway (b). Small-molecule inhibition of this dependency can then be exploited to selectively kill cancer cells.

(both DNA-repair proteins) which are sensitive to poly (ADP-ribose) polymerase (PARP) inhibition. This dependency appears to be due to these cell lines having an already damaged DNA repair apparatus and thus relying on PARP for their DNA repair.

Inhibition of PARP thus leads to these cells losing all DNA repair functions and ultimately causes death.[48] In line with this, the PARP inhibitor olaparib has shown promise in clinical trials of BRCA1/BRCA2-mutated patients with breast cancer.[49,50]

Cancer cell-line profiling

One problem with targeted cancer therapies is that currently only few patients actually benefit from them. Additionally, resistance is a commonly occurring problem.[51] In order to increase the number of patient-matched cancer therapies and to accelerate their discovery, cancer cell-line profiling has been used to identify novel dependencies. In these efforts, increasingly large numbers of cancer cell lines (CCLs) are tested for their sensitivity to small-molecule inhibition or inhibition of gene expression through RNA interference (RNAi). The sensitivities are then related to CCL features such as lineage, genetic, or epigenetic markers (Figure 1.3).[52-54] This approach has successfully confirmed the identification and subsequent targeting of dependencies on oncogenic

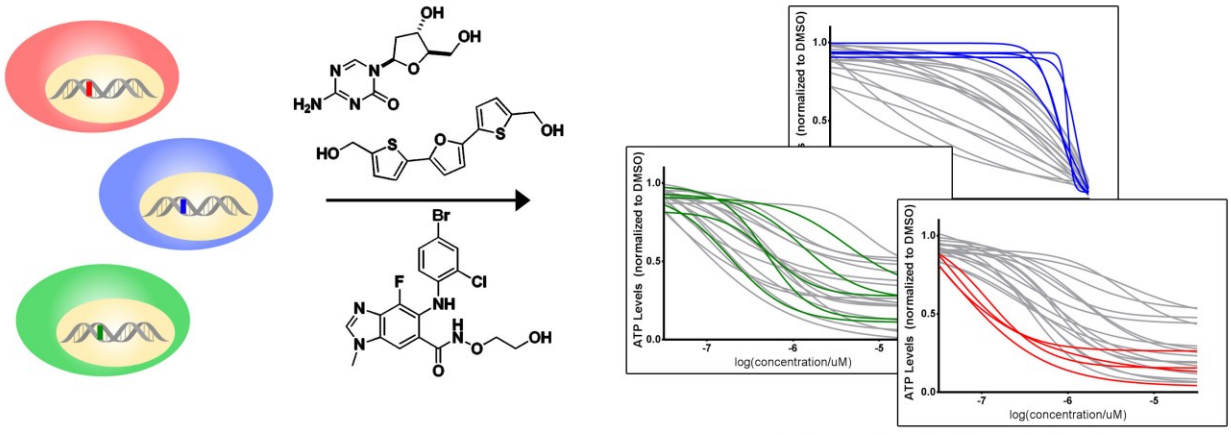


Figure 1.3. Cell-line profiling can help generate new hypotheses about cancer dependencies. Genetically and epigenetically characterized CCLs are tested for their sensitivity against a set of annotated small molecules. If many sensitive or non-responsive cell-lines share a feature, this might lead to a new hypothesis.

BRAF and *EGFR* alleles.[55] However, these studies are often limited by numbers of CCLs, compounds, or features characterized within the CCLs. New technologies and larger datasets allow more hypotheses to be generated and developed into therapies.

Roadmap

This thesis describes an effort to develop novel small-molecule probes and their application to cancer biology. In a pathway-targeted effort, a high-throughput screen for novel Hh pathway inhibitors and the classification of several known probes with additional activity as Smo inhibitors is described in Chapter Two. Chapter Three describes the identification of a group of molecules arising from the same screen as novel pathway inhibitors and the characterization of their distinct structure-activity relationships (SAR). Their mechanism of action is further elucidated in Chapter Four. In an unbiased approach, Chapter Five describes the generation of a large dataset of CCLs and their sensitivity to an annotated small-molecule informer set. Hypotheses generated from this approach, and follow-up experiments to one hypothesis, are then described in Chapter Six. Finally, Chapter Seven gives an outlook at future possibilities of small-molecule probe development in cancer biology.

References

(1) Camilli, A. & Bassler, B. L. "Bacterial small-molecule signaling pathways." *Science* **2006**, *311*, 1113-1116.

(2) Cooper, G. M. *The cell : a molecular approach*. 2nd edn, 523-532 (ASM Press; Sinauer Associates, **2000**).

(3) Lehninger, A. L., Nelson, D. L. & Cox, M. M. *Lehninger principles of biochemistry*. 4th edn, 421-479 (W.H. Freeman, **2005**).

(4) Walsh, C. "Molecular mechanisms that confer antibacterial drug resistance." *Nature* **2000**, *406*, 775-781.

(5) Scott, J. J. *et al.* "Bacterial protection of beetle-fungus mutualism." *Science* **2008**, *322*, 63.

(6) Schreiber, S. L. "Small molecules: the missing link in the central dogma." *Nat. Chem. Biol.* **2005**, *1*, 64-66.

(7) Schreiber, S. L. "Organic synthesis toward small-molecule probes and drugs." *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6699-6702.

(8) Borisy, G. G. & Taylor, E. W. "The mechanism of action of colchicine. Colchicine binding to sea urchin eggs and the mitotic apparatus." *J. Cell Biol.* **1967**, *34*, 535-548.

(9) Schiff, P. B., Fant, J. & Horwitz, S. B. "Promotion of microtubule assembly in vitro by taxol." *Nature* **1979**, *277*, 665-667.

(10) Belshaw, P. J. *et al.* "Synthesis, Structure and Mechanism in Immunophilin Research." *Synlett* **1994**, *1994*, 381-392.

(11) Miller, J. L. "Sirolimus approved with renal transplant indication." *Am. J. Health-Syst. Pharm.* **1999**, *56*, 2177-2178.

(12) Oprea, T. I. *et al.* "A crowdsourcing evaluation of the NIH chemical probes." *Nat. Chem. Biol.* **2009**, *5*, 441-447.

(13) Bunnage, M. E., Chekler, E. L. & Jones, L. H. "Target validation using chemical probes." *Nat. Chem. Biol.* **2013**, *9*, 195-199.

(14) An, W. F. & Tolliday, N. J. "Introduction: cell-based assays for high-throughput screening." *Methods Mol Biol (N.Y., NY, U.S.)* **2009**, *486*, 1-12.

(15) Bradner, J. E., McPherson, O. M. & Koehler, A. N. "A method for the covalent capture and screening of diverse small molecules in a microarray format." *Nat. Protoc.* **2006**, *1*, 2344-2352.

(16) Schenone, M., Dancik, V., Wagner, B. K. & Clemons, P. A. "Target identification and mechanism of action in chemical biology and drug discovery." *Nat. Chem. Biol.* **2013**, *9*, 232-240.

(17) Ziegler, S., Pries, V., Hedberg, C. & Waldmann, H. "Target identification for small bioactive molecules: finding the needle in the haystack." *Angew. Chem., Int. Ed.* **2013**, *52*, 2744-2792.

(18) Ingham, P. W. & McMahon, A. P. "Hedgehog signaling in animal development: paradigms and principles." *Genes Dev.* **2001**, *15*, 3059-3087.

(19) Chiang, C. *et al.* "Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function." *Nature* **1996**, *383*, 407-413.

(20) Roessler, E. *et al.* "Mutations in the human Sonic Hedgehog gene cause holoprosencephaly." *Nat. Genet.* **1996**, *14*, 357-360.

(21) Rubin, L. L. & de Sauvage, F. J. "Targeting the Hedgehog pathway in cancer." *Nat. Rev. Drug Discovery* **2006**, *5*, 1026-1033.

(22) Thayer, S. P. *et al.* "Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis." *Nature* **2003**, *425*, 851-856.

(23) Ng, J. M. & Curran, T. "The Hedgehog's tale: developing strategies for targeting cancer." *Nat. Rev. Cancer* **2011**, *11*, 493-501.

(24) Teglund, S. & Toftgård, R. "Hedgehog beyond medulloblastoma and basal cell carcinoma." *Biochim. Biophys. Acta* **2010**, *1805*, 181-208.

(25) Nüsslein-Volhard, C. & Wieschaus, E. "Mutations affecting segment number and polarity in Drosophila." *Nature* **1980**, *287*, 795-801.

(26) Peukert, S. & Miller-Moslin, K. "Small-molecule inhibitors of the hedgehog signaling pathway as cancer therapeutics." *ChemMedChem* **2010**, *5*, 500-512.

(27) Robbins, D. J., Fei, D. L. & Riobo, N. A. "The hedgehog signal transduction network." *Sci. Signaling* **2012**, *5*, re6.

(28) Taipale, J. *et al.* "Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine." *Nature* **2000**, *406*, 1005-1009.

(29) Robarge, K. D. *et al.* "GDC-0449-a potent inhibitor of the hedgehog pathway." *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5576-5581.

(30) Dlugosz, A., Agrawal, S. & Kirkpatrick, P. "Vismodegib." *Nat. Rev. Drug Discovery* **2012**, *11*, 437-438.

(31) Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. "Small molecule modulation of Smoothened activity." *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14071-14076.

(32) Sinha, S. & Chen, J. K. "Purmorphamine activates the Hedgehog pathway by targeting Smoothened." *Nat. Chem. Biol.* **2006**, *2*, 29-30.

(33) Stanton, B. Z. *et al.* "A small molecule that binds Hedgehog and blocks its signaling in human cells." *Nat. Chem. Biol.* **2009**, *5*, 154-156.

(34) Lauth, M., Bergström, A., Shimokawa, T. & Toftgård, R. "Inhibition of GLI-mediated transcription and tumor cell growth by small-molecule antagonists." *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8455-8460.

(35) Firestone, A. J. *et al.* "Small-molecule inhibitors of the AAA+ ATPase motor cytoplasmic dynein." *Nature* **2012**, *484*, 125-129.

(36) Bassilana, F. *et al.* "Target identification for a Hedgehog pathway inhibitor reveals the receptor GPR39." *Nat. Chem. Biol.* **2014**, *10*, 343-349.

(37) Jenkins, D. "Hedgehog signalling: emerging evidence for non-canonical pathways." *Cell. Signalling* **2009**, *21*, 1023-1034.

(38) Polizio, A. H., Chinchilla, P., Chen, X., Manning, D. R. & Riobo, N. A. "Sonic Hedgehog activates the GTPases Rac1 and RhoA in a Gli-independent manner through coupling of smoothened to Gi proteins." *Sci. Signaling* **2011**, *4*, pt7.

(39) Polizio, A. H. *et al.* "Heterotrimeric Gi proteins link Hedgehog signaling to activation of Rho small GTPases to promote fibroblast migration." *J. Biol. Chem.* **2011**, *286*, 19589-19596.

(40) Chinchilla, P., Xiao, L., Kazanietz, M. G. & Riobo, N. A. "Hedgehog proteins activate pro-angiogenic responses in endothelial cells through non-canonical signaling pathways." *Cell Cycle* **2010**, *9*, 570-579.

(41) Wang, Y., Zhou, Z., Walsh, C. T. & McMahon, A. P. "Selective translocation of intracellular Smoothened to the primary cilium in response to Hedgehog pathway modulation." *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2623-2628.

(42) Wang, Y. *et al.* "Selective identification of hedgehog pathway antagonists by direct analysis of smoothened ciliary translocation." *ACS Chem. Biol.* **2012**, *7*, 1040-1048.

(43) Epstein, E. H. "Basal cell carcinomas: attack of the hedgehog." *Nat. Rev. Cancer* **2008**, *8*, 743-754.

(44) Xie, J. *et al.* "Activating Smoothened mutations in sporadic basal-cell carcinoma." *Nature* **1998**, *391*, 90-92.

(45) Gailani, M. R. *et al.* "The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas." *Nat. Genet.* **1996**, *14*, 78-81.

(46) Luo, J., Solimini, N. L. & Elledge, S. J. "Principles of cancer therapy: oncogene and non-oncogene addiction." *Cell* **2009**, *136*, 823-837.

(47) Bollag, G. *et al.* "Vemurafenib: the first drug approved for BRAF-mutant cancer." *Nat. Rev. Drug Discovery* **2012**, *11*, 873-886.

(48) Savage, D. G. & Antman, K. H. "Imatinib mesylate--a new oral targeted therapy." *N. Engl. J. Med.* **2002**, *346*, 683-693.

(49) Haince, J. F., Rouleau, M., Hendzel, M. J., Masson, J. Y. & Poirier, G. G. "Targeting poly(ADP-ribosyl)ation: a promising approach in cancer therapy." *Trends Mol. Med.* **2005**, *11*, 456-463.

(50) Gelmon, K. A. *et al.* "Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study." *Lancet Oncol.* **2011**, *12*, 852-861.

(51) Gonzalez de Castro, D., Clarke, P. A., Al-Lazikani, B. & Workman, P. "Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance." *Clin. Pharmacol. Ther. (N.Y., NY, U.S.)* **2013**, *93*, 252-259.

(52) Shoemaker, R. H. "The NCI60 human tumour cell line anticancer drug screen." *Nat. Rev. Cancer* **2006**, *6*, 813-823.

(53) Garnett, M. J. *et al.* "Systematic identification of genomic markers of drug sensitivity in cancer cells." *Nature* **2012**, *483*, 570-575.

(54) Barretina, J. *et al.* "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity." *Nature* **2012**, *483*, 603-607.

(55) McDermott, U. *et al.* "Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling." *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19936-19941.

Page intentionally left blank.

Chapter Two

Effects of known small-molecule probes on the Hedgehog signaling pathway

Dr. Jose Perez and Dr. Benjamin Stanton developed and performed the cell-based Hedgehog screen and collected the conditioned medium used. Dr. Perez confirmed the hits in dose. I tested the hits' toxicity in Shh Light II cells, retested their effect in C3H10T1/2 cells, and prioritized candidates to further follow-up. I characterized these compounds' mechanisms of action in the SAG activation, $Ptch^{-/-}$, and BODIPY-cyclopamine assays.

In part adapted from Schaefer, G.I.; Perez, J.R.; Duvall, J.R.; Stanton, B.Z.; Shamji, A.F.; Schreiber, S.L. "Discovery of Small-Molecule Modulators of the Sonic Hedgehog Pathway" *J. Am. Chem. Soc.* **2013**, *135*, 9675-9680.[1]

Screen for inhibitors of Hedgehog pathway

In an effort to develop novel small-molecule probes for applications in cancer research, a pathway-targeted effort (Chapters 2 – 4) as well as an unbiased approach (Chapter 5 & 6) was employed. While inhibitors of the Hedgehog (Hh) pathway exist in the clinic, resistance is an issue and inhibitors with novel mechanisms of actions can help develop more potent therapies.[2] To enhance our molecular understanding of the Hh pathway, we aimed to discover novel small-molecule probes of Hh signaling. We first performed a cell-based high-throughput screen for novel inhibitors of Gli-induced transcription. 21,753 compounds were screened in a cell-based assay using Shh Light II cells. These cells are derived from NIH/3T3 cells by co-transfection with a Gli-responsive Firefly luciferase reporter.[3,4] All compounds were screened in duplicate at a single concentration. Screening positives (mean inhibition ≥ 65%) were re-tested in dose and their toxicity was assessed using CellTiter-Glo to measure cellular ATP levels as a surrogate for viability. 390 hits were identified and 180 were advanced for further investigation after toxicity measurements (Suppl. Figure S2.1, Suppl. Figure S2.2).

The compounds' activity on the Hh pathway was further verified using a second biological assay that measures alkaline phosphatase expression as a surrogate for Hh-induced differentiation of C3H10T1/2 cells into osteoblasts.[5] 160 compounds were confirmed in this assay and advanced as candidates for further investigation. Interestingly, when studying the confirmed hits, it was found that several of the compounds were annotated in the literature as having a mechanism of action other than inhibition of Hh signaling and the behavior of 11 of these compounds in pathway-relevant assays was further analyzed (Figure 2.1, Table 2.1).[1]

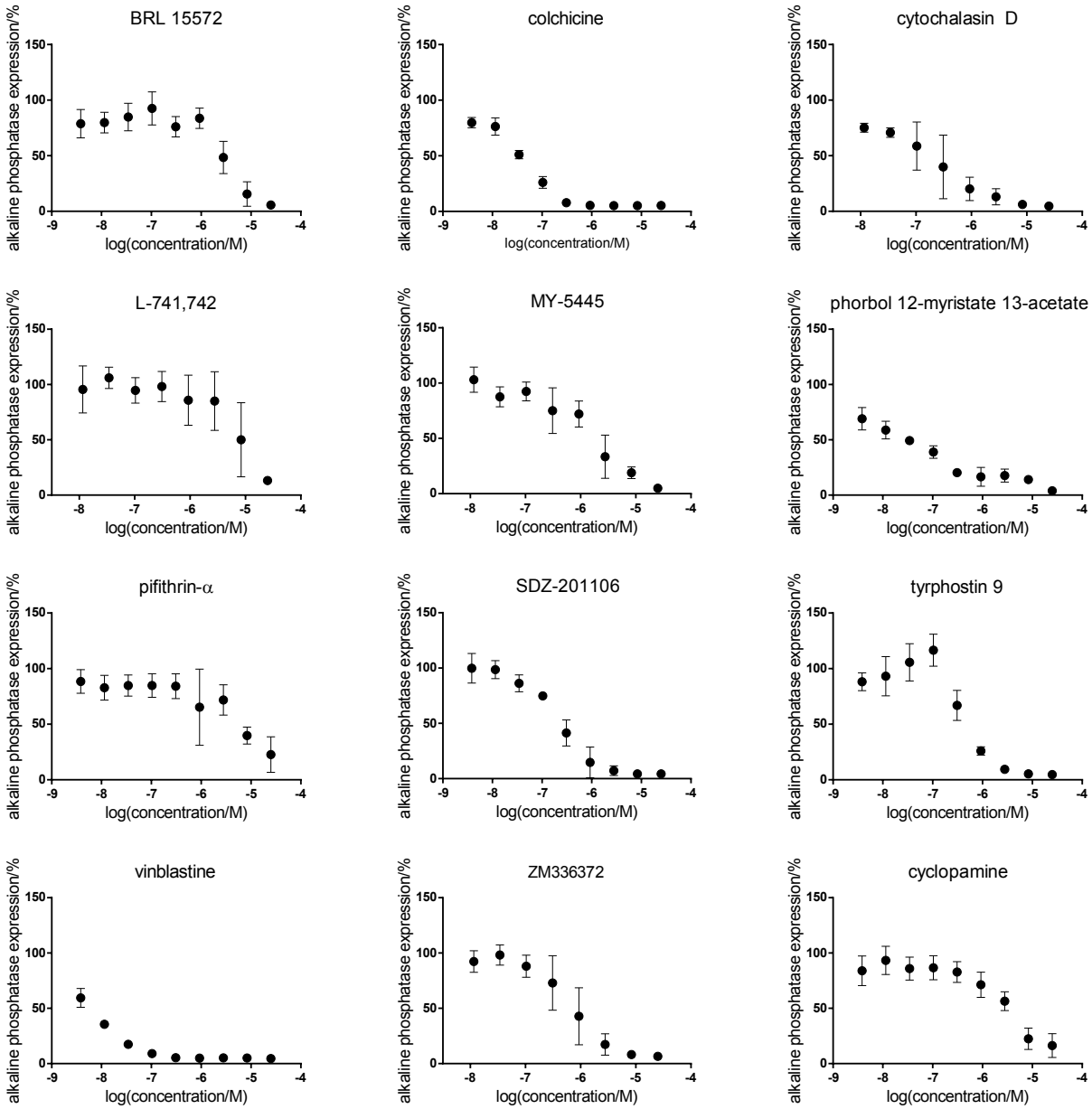


Figure 2.1. Inhibition of Shh-conditioned medium-induced differentiation of C3H10T1/2 cells by previously-annotated compounds, and cyclopamine as a control after 48h. All values are shown and generated from three (BRL 15572, colchicine, phorbol 12-myristate 13-acetate, pifithrin-a, SDZ-201106, tyrphostin 9, vinblastine, cyclopamine) or two (cytochalasin D, L-741,742, MY-5445, ZM336372) independent experiments run in duplicate (values are calculated average ± SD).

Table 2.1. list of candidates with previously annotated mechanisms of action

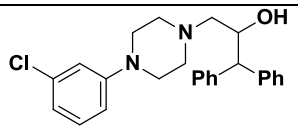
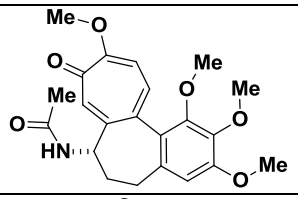
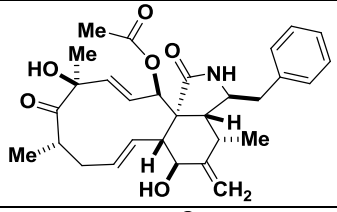
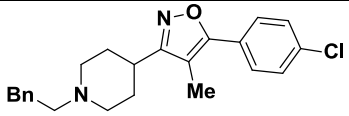
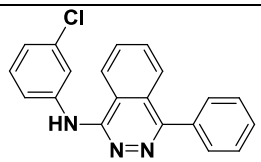
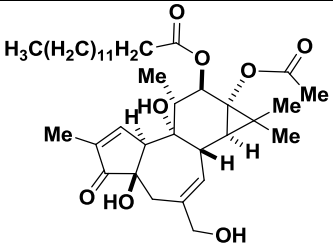
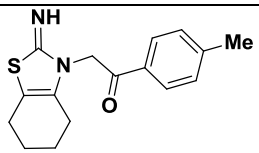
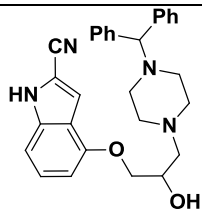
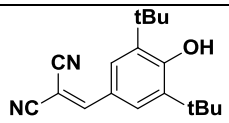
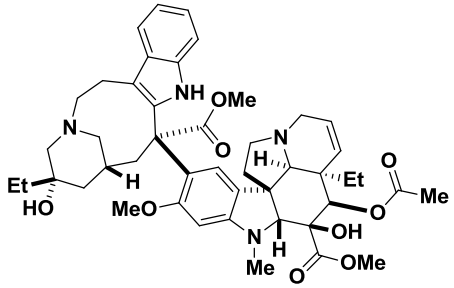
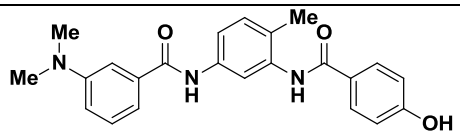
compound name	structure	selected annotated mechanism of action*
BRL 15572		selective h5-HT_{1D} antagonist[6]
colchicine		inhibitor of microtubule formation, binds tubulin[7]
cytochalasin D		inhibitor of actin polymerization[8]
L-741,742		antagonist of dopamine D4 receptor[9]
MY-5445		inhibitor of cGMP phosphodiesterase[10]
phorbol 12-myristate 13-acetate		activator of protein kinase C[11]
pifithrin-α		inhibitor of p53 induced apoptosis[12]
SDZ-201106		partial inhibitor of Na^+ pump, positive inotropic and calcium sensitizing agent[13,14]
tyrphostin 9		inhibitor of platelet-derived growth factor receptor tyrosine kinase[15]

Table 2.1. (Continued)

compound name	structure	selected annotated mechanism of action*
vinblastine		inhibitor of microtubule assembly[16]
ZM336372		*in vitro* inhibitor but *in vivo* activator of Raf-1[17]

*for compounds with several annotated mechanisms-of-action, representative examples were selected in this table.

Characterization of compounds' mechanism of action

The compounds' activity on the pathway could stem from different origins. It could be from on-target effects with their annotated target, resulting in cross-signaling or other perturbations of the Hh pathway. Alternatively, the compounds could have more than one cellular target and thus act on the Hh pathway directly. To gain more insight into the origins of the compounds' activity within the Hh pathway, additional experiments were conducted. The compounds were tested in Ptch$^{-/-}$ cells, mouse embryonic fibroblasts that contain a *β-galactosidase* reporter gene instead of the *Ptch* gene downstream of the *Ptch* promoter (Figure 2.2).[3] The pathway in these cells is constitutively activated through the loss of *Ptch*, and compounds only retain activity in this assay if they act at a step downstream or at Ptch, but not upstream of it. Additionally, the compounds were assayed in C3H10T1/2 cells where the pathway is activated with SAG (a small-molecule activator of Smo) rather than Shh-conditioned media (Figure 2.3).[4] Activity of compounds in this assay suggests that the compounds

act at or below of Smo in the pathway. The compounds were active in both assays, suggesting that they act at a step at or below Smo in the pathway. Some compounds show toxicity in these assays, but less than inhibition at the respective concentrations.

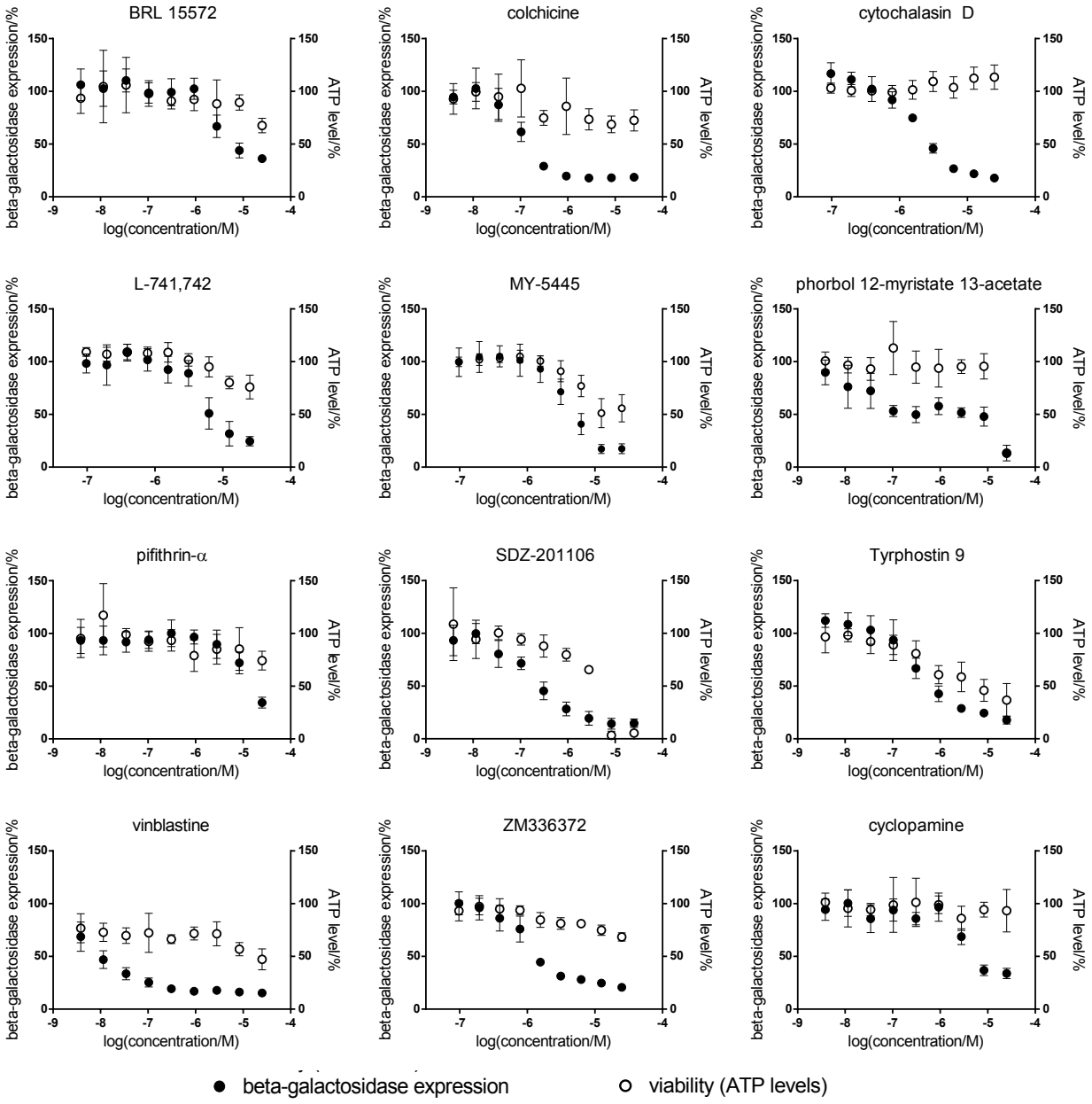


Figure 2.2. *β-galactosidase* expression (black dots) and viability (white dots) (measured using ATP levels as a surrogate) response of $Ptch^{-/-}$ cells to 48 h treatment with previously annotated compounds and cyclopamine as a control. All values are shown and generated from three independent experiments run in duplicate (values are calculated average ± SD).

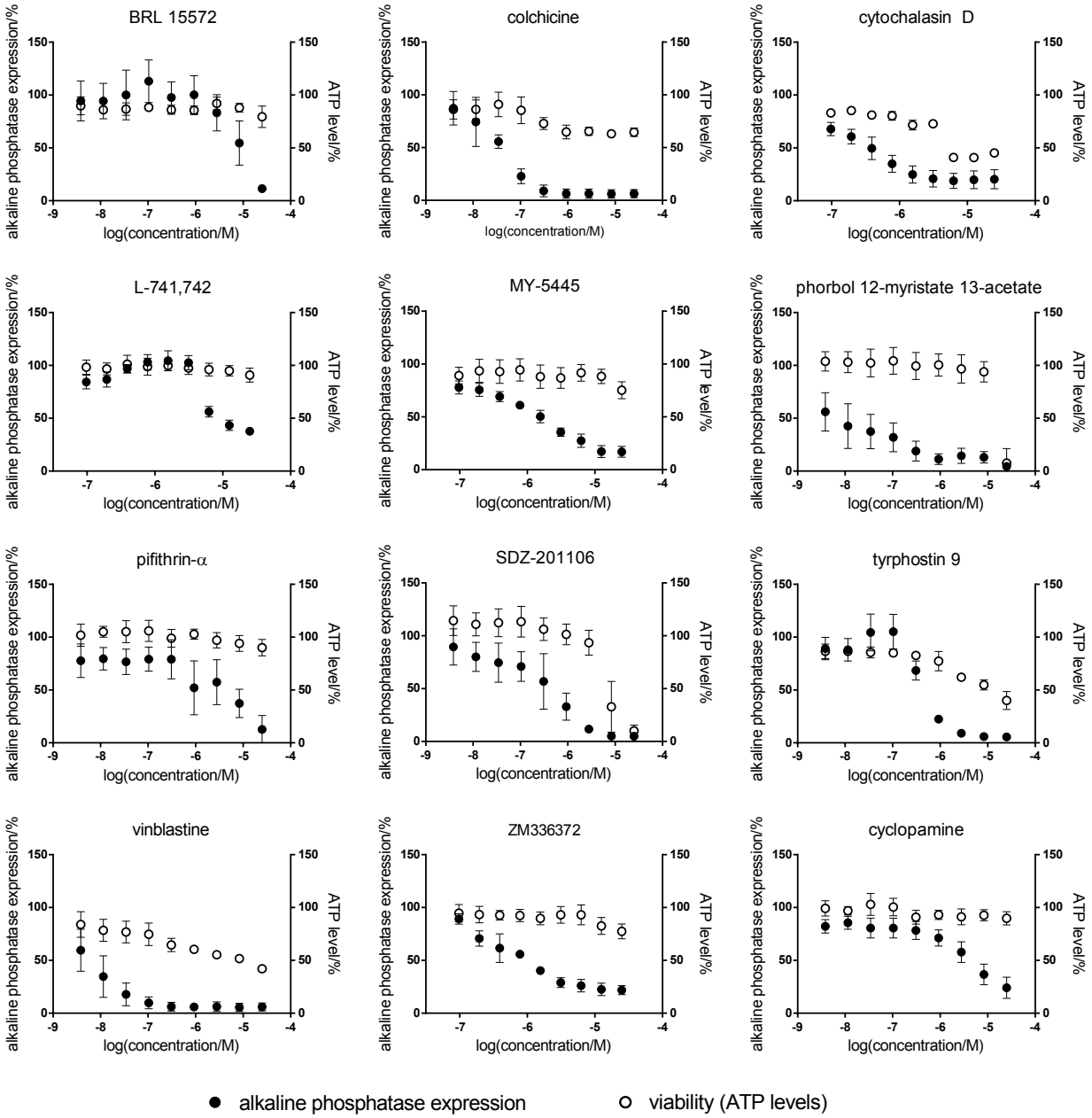


Figure 2.3. Inhibition of SAG-induced differentiation (black dots) and viability (white dots) (measured using ATP levels as a surrogate) of C3H10T1/2 cells by previously annotated compounds and cyclopamine as a control after 48h. All values are shown and generated from three independent experiments run in duplicate (values are calculated average ± SD).

To directly test the compounds' ability to bind Smo, a protein known to interact with many molecules, the compounds were tested in a competition assay. In this cellular assay, it is determined whether compounds displace BODIPY-cyclopamine, a

fluorescent cyclopamine-derivative. The HEK293T cells are transfected with a smo-myc-his construct, which can also be probed for using an anti-myc antibody. Loss of BODIPY-cyclopamine signal in transfected cells thus suggests that molecules bind Smo in the cyclopamine-binding site (Figure 2.4).[18]

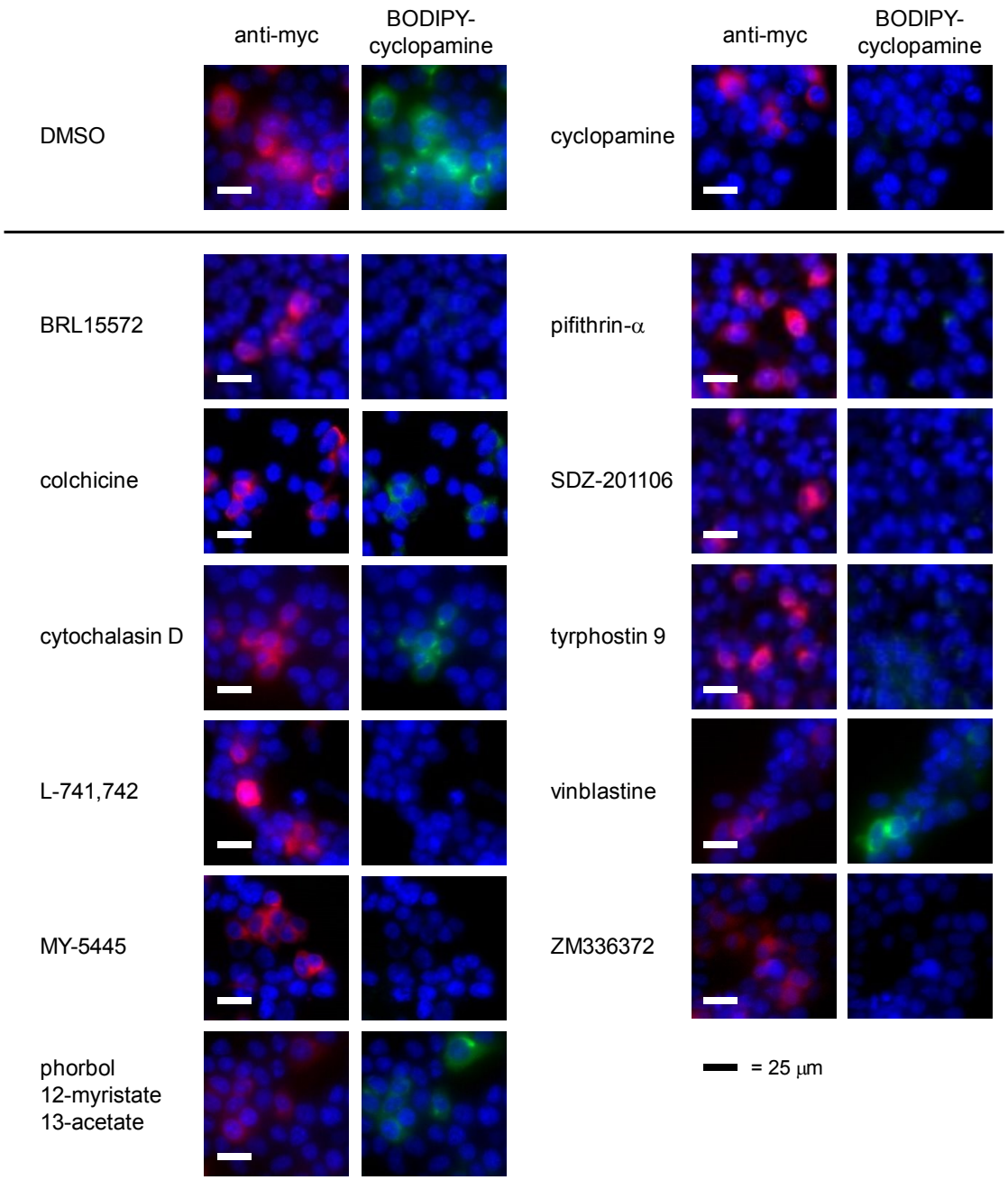


Figure 2.4. Effects of DMSO, cyclopamine, and the compounds (10 μM) on BODIPY-cyclopamine (5-10 nM) binding to exogenously expressed Smo. blue (Hoechst 33342), red (anti-myc), green (BODIPY-cyclopamine).

Several annotated probes appear to interact with Smo

BRL 15572, L-741,742, MY-5445, pifithrin-α, SDZ-201106, tyrphostin 9, and ZM336372 displaced BODIPY-cyclopamine in this assay, leading to the conclusion that Smo is one of their cellular targets. For these compounds, the activity on the Hh pathway most likely stems from Smo inhibition. These compounds have previously been annotated to interact with different cellular targets, some of them selectively over other targets. While cross-signaling with the proteins from these annotated interactions might be partially responsible for the Hh pathway perturbation, the displacement of BODIPY-cyclopamine suggests an interaction of the compounds with Smo and thus a potential additional target for these compounds (Figure 2.4).

For the compounds that did not displace BODIPY-cyclopamine (colchicine, cytochalasin D, vinblastine, and phorbol 12-myristate 13-acetate), the action on the pathway cannot be clearly classified as being due to cross-signaling with their annotated target or as being an off-target effect from their original annotation. Colchicine and vinblastine are anti-microtubule agents and cytochalasin D is a disruptor of actin filament formation. Both processes have been shown to affect the Hh pathway by influencing the primary cilium and protein trafficking.[19-21] Activation of protein kinase c using phorbol 12-myristate 13-acetate has also been shown to negatively regulate the Hh pathway.[22] The mechanism of action of these compounds is thus most likely on-target cross-signaling into the pathway.

The data presented above thus suggests that caution should be exercised with some of the previously reported probes. Their bioactivity might not only stem from their annotated mechanisms of action, which makes it harder to interpret their activity in

different bio-assays. When designing an experiment, the concentrations at which these compounds display activity with their originally annotated targets should be taken into account (Table 2.2).

Table 2.2. Concentrations at which previously annotated compounds display activity

compound	concen-tration	activity at concentration	% Hh inhibition[a]
BRL 15572	0.008 μM	EC_{50} of h5-HT_{1D} receptor expressed in CHO cells (0.75 intrinsic activity)[6]	21.8
L-741,742	1 μM	antagonizes inhibition mediated by 1μM dopamine of 10 μM forskolin-induced elevation of cAMP levels in HEK293 cells[9]	14.2
MY-5445	0.6 μM	IC_{50} of purified cyclic GMP phosphodiesterase (FI)[10]	27.9
pifithrin-α	10 μM	inhibits apoptotic cell death induced by doxorubicin, etoposide, taxol, or cytosine arabinose in C8 cells[12]	60.2
SDZ-201106	1 μM	IC_{50} of $^{22}Na^{+}$ uptake by chicken cardiac cells[14]	85.3
tyrphostin 9	0.04 μM	IC_{50} of PDGF induced DNA synthesis in RVMSC cells[15]	-5.7
ZM336372	0.07 μM	IC_{50} of c-Raf in cell lysates from v-Ras and Lck-transfected Sf9 cells[17]	1.8

[a]% inhibition of CM-induced osteoblast differentiation at dose point closest to concentration reported in literature

The concentration of some compounds' interaction with their target is significantly below that of the interactions with the Hh pathway and does not cause significant % inhibition of osteoblast differentiation. One example of this is BRL15572, which was annotated as an inhibitor of the h5-HT_{1D} receptor with more than 100 fold selectivity over h5-HT_{1B}.[6] Therefore, if this compound is used at low concentrations, and the main reason for its use is to distinguish between the two receptors, using it in a study should minimize influences from the Smo off-target effect.

The relevant concentration for MY-5445 lies significantly below that at which it shows activity in the Hh pathway. However, that activity was not measured in cells and its cellular activity might be at higher concentrations. For L-741,742 and SDZ-201106, the concentrations that were used in their original studies are significantly closer to its active concentration in the Hh pathway. For studies using these compounds, caution must be exercised when interpreting their biological effects. In order to confirm findings made using these compounds, different probes that hit the desired target, ideally with distinct structural features, can be used.

While the IC_{50} from the literature for ZM336372 is also significantly below the concentration that the Hh inhibition is seen at, this compound has given an interesting response. *In vitro*, it has been shown to inhibit c-Raf, while its *in vivo* activity, contrary to that, actually activates c-Raf. This could be due to a feed-back loop, or due to off-target effects. One off-target that has been identified is SAPK/p38.[17] Therefore, this compound has a confusing profile, with Smo targeting being an additional factor and interpretations of its activity should be made carefully.

The actual target for pifithrin-α has not yet been identified, but it is used as an inhibitor of p53-mediated apoptosis.[12] It has been shown that a feedback loop exists between the Hh signaling pathway and p53.[23,24] Inhibition of Smo could thus be involved in pifithrin-α's mechanism of action. Further experiments would be needed to confirm this hypothesis.

Conclusion

In conclusion, the screen found several compounds that had previously annotated as modulators to other biological pathways. I show that the compounds also inhibit the Hh pathway, and appear to target Smo. When using these compounds in experiments, caution should be exercised in interpreting the data based on just the annotated target, as Smo inhibition might also be a factor. In some cases, this can be avoided by using the compounds at lower concentrations. However, the large number of alternately-annotated compounds that target Smo found in this screen suggests that many more compounds have multiple mechanisms of action. If looking at other compounds or receptors, one might find more off-target effects of many different probes. Therefore, having structurally distinct, selective probes targeting the same protein and probes which target different proteins of a pathway, is an advantage when using tool molecules to dissect biological findings and pathways. The goal of the chemical biology community should be to fill in the gaps in targets with structurally diverse small-molecule probes.

Experimental Section

Materials and Methods

Screening: Compounds screened were taken from the commercial, natural products, bioactives, and diversity-oriented synthesis (DOS) libraries available at the Broad Institute (Cambridge, MA, USA).

Cell Lines and Reagents: HEK293T, C3H10T1/2, and Shh Light II cells were purchased from ATCC (Manassas, VA, USA). $Ptch^{-/-}$ cells were kindly provided by Dr.

James Chen (Stanford University, CA, USA). Shh Light II cells were grown at 37°C/5% CO_2 in DMEM supplemented with 10% bovine calf serum (BCS) (ATCC, VA, USA), 1% penicillin/streptomicyn (Cellgro, VA, USA), 2 mM L-glutamine (Life Technologies, Carlsbad, CA, USA), 0.4 mg/ml G-418 (EMD Millipore Chemicals, Billerica, MA, USA), and 0.15 mg/ml Zeocin (Invitrogen, Carlsbad, CA, USA). All other cell lines were maintained at 37°C/5% CO_2 in DMEM supplemented with 10% fetal bovine serum (FBS) (Invitrogen, Carlsbad, CA, USA), and 1% penicillin/streptomicyn. Shh Light II cells were activated using a 1:25 dilution of Shh-conditioned medium in LCCM (phenol red-free DMEM (Cellgro, Manassan, VA, USA) with 0.5% bovine calf serum, 1% penicillin/streptomycin, and 800 μM L-Gln prepared the same day). The differentiation of C3H10T1/2 cells was induced by 4-5% Shh conditioned medium or 25-30 nM SAG in DMEM (depending on EC_{80} of induction determined in a pilot experiment for each freshly thawed batch of cells) supplemented with 0.5% FBS and 1% penicillin/streptomycin. Shh conditioned medium was previously collected from Shh-N overexpressing HEK 293T cells (kindly provided by Dr. James Chen (Stanford University, USA)) as reported before.[4]

Assay Treatments and Readouts: Small-molecule controls were purchased from Sigma-Aldrich (St Louis, MO, USA), Alfa-Aesar (Ward Hill, MA, USA), Enzo Life Sciences (Farmingdale, NY, USA), Toronto Research Chemicals (North York, ON, Canada), and EMD Millipore Chemicals (Billerica, MA, USA). Alkaline-Phosphatase buffer was purchased from Boston Bioproducts (Ashland, MA, USA), and CDP-Star® Western Blot Chemiluminescence Reagent was purchased from Perkin Elmer (Waltham, MA, USA). Bright-Glo, Beta-Glo, and CellTiter-Glo reagents were purchased from

Promega (Madison, WI, USA). Antibodies were purchased from Life Technologies (Carlsbad, CA, USA), and Roche Diagnostics (Indianapolis, IN, USA). BODIPY-cyclopamine was purchased from Medical Isotopes (Pelham, NH, USA) and Toronto Research Chemicals (North York, ON, Canada). Smo-Myc-His construct was kindly provided by Dr. James Chen (Stanford University, CA, USA). Plate Reader settings used were: luminescence settings: Perkin Elmer (Waltham, MA, USA) Envision 2104 Multilabel Plate Reader, luminescence mirror, luminescence 700 emission filter, 0.1 s measurement time, 6.5 mm measurement height; US luminescence settings: Perkin Elmer Envision 2102 or 2104 Multilabel Plate Reader, 384 plate US luminescence aperture, 0.5 mm distance between plate and detector, 0.5 s measurement time, 0% Glow (CT2) correction factor. High-content microscope images were taken on a Molecular Devices ImageXpress High-Content Microscope (Sunnyvale, CA, USA) using DAPI settings for nuclear staining (10-80 ms exposure), GFP settings for BODIPY-cyclopamine (160 ms-1 s exposure), and Texas Red settings for Alexa Fluor 568 antibody (80-160 ms exposure), or on a Molecular Devices ImageXpress 5000 High-Content Microscope using DAPI settings for nuclear staining (3 ms exposure), FITC settings for BODIPY-cyclopamine (80 ms exposure), and Rhodamine settings for Alexa Fluor 568 antibody (10 ms exposure).

Assay Protocols

Shh Light II screen/assay: To measure Hh activity in high-throuput, a luciferase-based assay was performed using Shh Light II cells with a Bright-Glo luciferase reagent readout. Shh Light II cells were seeded in clear bottom, white 384 well plates at 3,500

cells/well density in 50 μL/well DMEM supplemented with 10% BCS, 2 mM L-glutamine, 1% penicillin/streptomicyn, 0.4 mg/ml G-418, and 0.15 mg/ml zeocin. Plates were covered with rayon plate covers and incubated for 4 days at 37°C/5% CO_2. After removal of the covers, the medium was removed from the plates and they were washed with 50 μL/well PBS. The PBS was removed and 40 μL/well of a 1:25 dilution of Shh-conditioned medium in LCCM (phenol red-free DMEM with 0.5% bovine calf serum, 1% penicillin/streptomycin, and 800 μM L-Gln prepared the same day) was added. The compounds were pinned with 100 nL/well, new rayon plate covers were applied, and the plates were incubated for 26-30 hours. The plate covers were removed and the medium aspirated. The plates were covered with a foil seal and frozen at -80°C for 1 hour. Afterwards, the plates were thawed at room temperature for 10 min, at 37°C for 15 min and at room temperature for another 10 min to equilibrate. 50 μL/well Bright-Glo luciferase reagent was added. The plates were shaken at room temperature for 2 min, and if there were bubbles, they were carefully blown with a heat gun to pop the bubbles (but not to heat the liquid). The plates were read out using US luminescence settings.

Shh Light II CellTiter-Glo Viability assay: To control for toxicity leading to the effects seen, a viability assay was performed in the Shh Light II cells. Shh Light II cells were seeded in clear bottom, white 384 well plates at 2,500 cells/well density in 50 μL/well DMEM supplemented with 10% BCS, 2 mM L-glutamine, 1% penicillin/streptomicyn, 0.4 mg/ml G-418, and 0.15 mg/ml zeocin. The cells were incubated at room temperature for 1 h and subsequently for 1 day at 37°C/5% CO_2. The cells were treated as in the Shh Light II assay above, but after pinning of the compounds, the plates were pinned again with a staurosporine control plate containing

1 mM staurosporine and DMSO. The plates were subsequently incubated at 37°C/5% CO_2 for 30 hours. After the incubation period, the plates were removed from the incubator and allowed to come to room temperature. The plate covers were removed, and the medium aspirated. Subsequently, 20 μL of phenol-red free DMEM was added to each well, followed by 20 μL of a 1:3 dilution of CellTiter-Glo reagent in PBS and the plates were shaken at room temperature for two minutes. The plates were then incubated at room temperature for at least 10 min and read out using luminescence settings.

C3H10T1/2 differentiation assay: To validate the compounds in a different biological readout, the C3H10T1/2 differentiation assay was performed. C3H10T1/2 cells were seeded in solid bottom, white 384 well plates at 3000 cells/well density in 50 μL/well DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Cells were incubated at 37°C/5% CO_2 until confluent (approximately 36 hours). The medium was removed and each well was washed with 40 μL PBS. 40 μL of DMEM supplemented with 0.5% FBS, 1% penicillin/streptomycin, and either 4-5% Shh conditioned medium, or 25-30 nM SAG was added to each well. Subsequently, the compounds were pinned at their respective concentrations with 100 nL/well. The plates were incubated at 37°C/5% CO_2 for 48 hours. After the incubation period was over, the medium was removed and 20 μL of lysis buffer (Alkaline-Phosphatase-1 buffer with 1% Triton-X 100) was added to each well. The plates were shaken for 45 min at room temperature. Subsequently, 50 μL of CDP-Star® Western Blot Chemiluminescence Reagent (for the detection of alkaline phosphatase) was added to each well and the plates were incubated for 30 min before being read out using US luminescence settings.

Ptch$^{-/-}$ β-galactosidase assay: To test the compounds' mechanism of action, they were tested in Ptch$^{-/-}$ mouse embryonic fibroblasts. Ptch$^{-/-}$ cells were seeded in solid bottom, white 384 well plates at 1000 cells/well density in 50 μL/well DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Cells were incubated at 37°C/5% CO_2 for 48 hours. The medium was removed and each well was washed with 40 μL PBS. 40 μL of DMEM supplemented with 0.5% FBS, 1% penicillin/streptomycin was added to each well. Subsequently, the compounds were pinned at their respective concentrations with 100 nL/well. The plates were incubated at 37°C/5% CO_2 for 48 hours. After the incubation period was over, the plates were removed from the incubator and allowed to equilibrate to room temperature. Subsequently, 20 μL of Beta-Glo luminescence reagent was added to each well and the plates were shaken at room temperature for 30 min before being read out using luminescence settings.

C3H10T1/2 and Ptch$^{-/-}$ CellTiter-Glo Viability assay: The cells were plated as in the C3H10T1/2 differentiation assay or the Ptch$^{-/-}$ beta-galactosidase assay above. The cells were then incubated as in the assays above. The cells were treated as in the C3H10T1/2 differentiation assay or the Ptch$^{-/-}$ beta-galactosidase assay above, but after pinning of the compounds, the plates were pinned again with a staurosporine control plate containing 1 mM staurosporine and DMSO. The plates were subsequently incubated at 37°C/5% CO_2 for 48 hours. After the incubation period was over, the plates were removed from the incubator and allowed to come to room temperature. Subsequently, 20 μL of a 1:1 dilution of CellTiter-Glo reagent in PBS was added to each well and the plates were shaken at room temperature for a few seconds. The plates

were then incubated at room temperature for at least 10 min and read out using luminescence settings.

BODIPY-cyclopamine competition assay: To test the compounds' ability to compete with BODIPY-cyclopamine binding, they were tested in a BODIPY-cyclopamine imaging assay. HEK293T cells were plated in clear-bottom, black 96-well plates, 10,000 cells/well, 100 μL/well in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin pre-coated with poly-D-lysine. The cells were incubated at 37°C/5% CO_2 for 48 hours. The cells were then transfected with a Smo-Myc-His construct (kindly provided by Dr. James Chen (Stanford University, USA)) using Fugene 6 transfection reagent. The plates were incubated for another 24 hours after which they were carefully washed with PBS, and 100 μL DMEM supplemented with 0.5% FBS and 1% penicillin/streptomycin preincubated with the compounds (10 μM) and BODIPY-cyclopamine (10 nM) was added to each well. The plates were then incubated for 30 min at 37°C/5% CO_2. Afterwards, each well was washed twice with 100 μL PBS, and 100 μL formaldehyde (diluted 1:10 in PBS) was added to each well to fix the cells. After 30 min of incubation at room temperature, each well was washed three times with 100 μL PBS, and 100 μL of 1% FBS in PBS was added to block the cells. After another 60 min of incubation at room temperature, the FBS solution was removed and 100 μL 1° antibody (anti-c-myc from mouse IgG, 9E10, 1:1000 dilution in PBS) was added to each well. The plates were subsequently incubated for 4 hours at room temperature. Afterwards, each well was washed twice with 100 μL PBS, and 2° antibody (Alexa Fluor 568 donkey anti-mouse, 1:1000 in PBS) was added to each well. After 60 min of incubation at room temperature, each well was washed twice with 100 μL PBS, and 100

μL of 10 μM Hoechst 33342 in PBS was added to each well and the plates were incubated for 10-15 min at room temperature. Finally, the plates were washed twice with 100 μL/well PBS, and 100 μL PBS were added to each well before the plates were imaged. Keeping the plates in the dark during the incubation periods improves the quality of the images.

Data Analysis

Primary screening, dose response, and viability data (Shh Light II cells): Primary screening and dose response data was normalized to neutral (NC; n = 36) and positive control (PC; n = 36) wells present on each plate using the Genedata Assay Analyzer (v7.3). Active compounds in the primary screen were identified as compounds that inhibited the *Gli1* luciferase reporter expression by >65% relative to the positive control cyclopamine at 12.5 μM. For EC_{50} determinations of screening positives in the same assay data points were also normalized to neutral (0%) and positive control (-100%) wells and curve fits were performed using the fitting models in the Genedata Screener Condoseo package (7.0.3). Viability data was normalized to neutral (0%) and positive control (-100%, staurosporine) wells.

Dose response curves (C310T1/2 and Ptch$^{-/-}$ cells): Mock treatments for each plate were averaged and used to calculate a percent response for each compound treatment on the corresponding plate. Percent responses were averaged for each dose point (every assay was performed two or three times in duplicate on separate days). Error bars are based on the standard deviation of the percent responses of each

concentration. % inhibitory data was calculated by subtracting the corresponding activity value from 100%

References

(1) Reprinted (adapted) with permission from Schaefer, G. I. *et al.* "Discovery of Small-Molecule Modulators of the Sonic Hedgehog Pathway." *J. Am. Chem. Soc.*, *135*, 9675-9680. Copyright (**2013**) American Chemical Society.

(2) Metcalfe, C. & de Sauvage, F. J. "Hedgehog fights back: mechanisms of acquired resistance against Smoothened antagonists." *Cancer Res.* **2011**, *71*, 5057-5061.

(3) Taipale, J. *et al.* "Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine." *Nature* **2000**, *406*, 1005-1009.

(4) Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. "Small molecule modulation of Smoothened activity." *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14071-14076.

(5) Spinella-Jaegle, S. *et al.* "Sonic hedgehog increases the commitment of pluripotent mesenchymal cells into the osteoblastic lineage and abolishes adipocytic differentiation." *J. Cell Sci.* **2001**, *114*, 2085-2094.

(6) Price, G. W. *et al.* "SB-216641 and BRL-15572--compounds to pharmacologically discriminate h5-HT1B and h5-HT1D receptors." *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1997**, *356*, 312-320.

(7) Hastie, S. B. "Interactions of colchicine with tubulin." *Pharmacol. Ther.* **1991**, *51*, 377-401.

(8) Carlier, M. F., Criquet, P., Pantaloni, D. & Korn, E. D. "Interaction of cytochalasin D with actin filaments in the presence of ADP and ATP." *J. Biol. Chem.* **1986**, *261*, 2041-2050.

(9) Rowley, M. *et al.* "5-(4-Chlorophenyl)-4-methyl-3-(1-(2-phenylethyl)piperidin-4-yl)isoxazole: a potent, selective antagonist at human cloned dopamine D4 receptors." *J. Med. Chem.* **1996**, *39*, 1943-1945.

(10) Hagiwara, M., Endo, T., Kanayama, T. & Hidaka, H. "Effect of 1-(3-chloroanilino)-4-phenylphthalazine (MY-5445), a specific inhibitor of cyclic GMP phosphodiesterase, on human platelet aggregation." *J. Pharmacol. Exp. Ther.* **1984**, *228*, 467-471.

(11) Castagna, M. *et al.* "Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor-promoting phorbol esters." *J. Biol. Chem.* **1982**, *257*, 7847-7851.

(12) Komarov, P. G. *et al.* "A chemical inhibitor of p53 that protects mice from the side effects of cancer therapy." *Science* **1999**, *285*, 1733-1737.

(13) Kaumann, A. J., Richards, D. E. & Russell, D. A. "Inhibition of the sodium pump by cardioactive DPI 201-106." *Br. J. Pharmacol.* **1987**, *91*, 3-5.

(14) Romey, G. *et al.* "Na+ channels as sites of action of the cardioactive agent DPI 201-106 with agonist and antagonist enantiomers." *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 896-900.

(15) Bilder, G. E. *et al.* "Tyrphostins inhibit PDGF-induced DNA synthesis and associated early events in smooth muscle cells." *Am. J. Physiol.* **1991**, *260*, C721-730.

(16) Na, G. C. & Timasheff, S. N. "Thermodynamic linkage between tubulin self-association and the binding of vinblastine." *Biochemistry* **1980**, *19*, 1355-1365.

(17) Hall-Jackson, C. A. *et al.* "Paradoxical activation of Raf by a novel Raf inhibitor." *Chem. Biol. (Oxford, U.K.)* **1999**, *6*, 559-568.

(18) Sinha, S. & Chen, J. K. "Purmorphamine activates the Hedgehog pathway by targeting Smoothened." *Nat. Chem. Biol.* **2006**, *2*, 29-30.

(19) Zhu, A. J., Zheng, L., Suyama, K. & Scott, M. P. "Altered localization of Drosophila Smoothened protein activates Hedgehog signal transduction." *Genes Dev.* **2003**, *17*, 1240-1252.

(20) Firestone, A. J. *et al.* "Small-molecule inhibitors of the AAA+ ATPase motor cytoplasmic dynein." *Nature* **2012**, *484*, 125-129.

(21) Sanders, T. A., Llagostera, E. & Barna, M. "Specialized filopodia direct long-range transport of SHH during vertebrate tissue patterning." *Nature* **2013**, *497*, 628-632.

(22) Cai, Q., Li, J., Gao, T., Xie, J. & Evers, B. M. "Protein kinase Cdelta negatively regulates hedgehog signaling by inhibition of Gli1 activity." *J. Biol. Chem.* **2009**, *284*, 2150-2158.

(23) Po, A. *et al.* "Hedgehog controls neural stem cells through p53-independent regulation of Nanog." *EMBO J.* **2010**, *29*, 2646-2658.

(24) Zbinden, M. *et al.* "NANOG regulates glioma stem cells and is essential in vivo acting in a cross-functional network with GLI1 and p53." *EMBO J.* **2010**, *29*, 2659-2674.

Chapter Three

Discovery and optimization of small-molecule modulators of the Sonic Hedgehog pathway

Dr. Jose Perez and Dr. Benjamin Stanton developed and performed the cell-based hedgehog screen presented in this chapter and collected the conditioned medium used. Dr. Perez confirmed the hits in dose. I tested the hits' toxicity in Shh Light II cells and retested the effect of all stereoisomers and analogs in C3H10T1/2 cells. I synthesized all additional analogs, characterized them, and measured their response in cells. The analytical platform at the Broad Institute measured the solubility of the compounds.

In part adapted from Schaefer, G.I.; Perez, J.R.; Duvall, J.R.; Stanton, B.Z.; Shamji, A.F.; Schreiber, S.L. "Discovery of Small-Molecule Modulators of the Sonic Hedgehog Pathway" *J. Am. Chem. Soc.* **2013**, *135*, 9675-9680.[1]

Confirmation of additional screening hits

The screen presented in the previous chapter had additional results of great interest. Both the primary screen data and multiple dose-retest data revealed a striking correlation between activity and stereochemistry of members of a library of the screening collection. These compounds were initially synthesized using the build/couple/pair strategy of diversity-oriented synthesis (DOS).[2,3] As a consequence, all possible stereoisomers of each structural type are included in the collection. The compounds in the library screened include approximately 6,700 compounds with varying eight-membered rings that are formed by nucleophilic aromatic substitution reactions. Based on the primary screening data, two of the eight stereoisomers of several compounds having the same eight-membered ring skeleton were active, the RSR and the SSR isomer, with the sole difference being the configuration of the extra-annular methyl group (Figure 3.1). These initial findings were confirmed in the second biological assay using Hh-induced differentiation of C3H10T1/2 cells. Re-testing all eight stereoisomers of several hit compounds in dose using Shh-conditioned medium-induced C3H10T1/2 cells confirmed that BRD50837, the RSR isomer, selectively blocks the Hh pathway in cells (Figure 3.1b, Suppl. Figure S3.1).

All hits were tested in the secondary differentiation assay using C3H10T1/2 cells to substantiate their on-pathway activity (Table 3.1, Suppl. Table S3.1). Additionally, to rule out gross toxicity as source of signal, all compounds were tested in a viability assay using CellTiter-Glo as a means to estimate cellular levels of ATP (Suppl. Table S3.1). BRD50837 displayed high potency with an EC_{50} of 0.09 μM. A PubChem search of other assays wherein BRD50837 (CID 44499307) was screened revealed that, as of

May 27, 2013, BRD50837 had been tested in 31 different assays, but only scored in our initial screen, suggesting that it is not broadly active. Compared to other similarly potent compounds, BRD50837 showed good PBS solubility (64.3 μM), and was thus chosen as a starting point for further experimentation (Table 3.1).

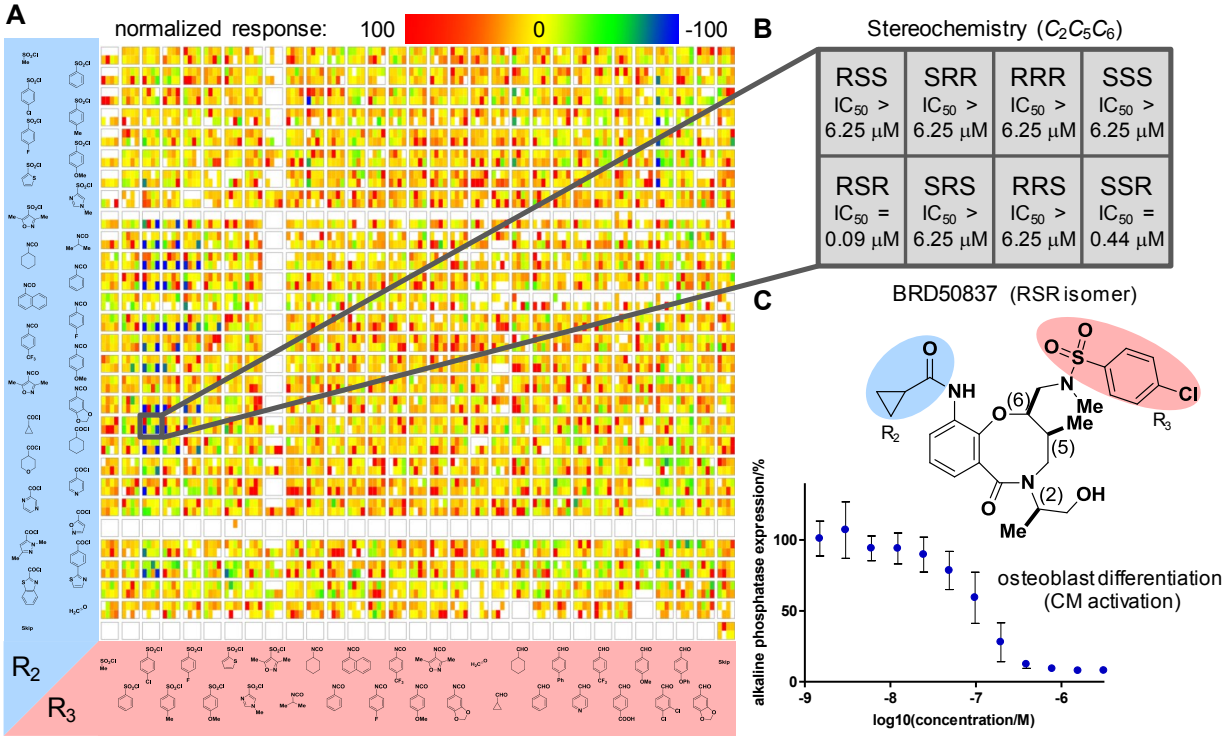
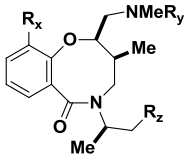
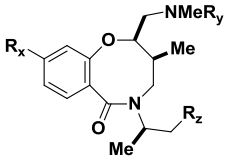


Figure 3.1. a) Primary screening data are displayed as percent luciferase activity in Shh Light II cells. Each small block displays a heat map associated with eight stereoisomeric compounds having the same skeleton and appendages. The overall panel A displays a near-complete matrix of multiple skeletons (only one shown in C) and building blocks used for R_2 and R_3 (displayed on the left (y-axis) and bottom (x-axis)). In the heat maps, blue represents inhibition (-100), yellow represents no activity (0), and red represents activation (100) normalized to DMSO control. Empty cells represent compounds not tested. Values shown are from testing compounds twice in single dose. b) The highlighted block represents the dose response data of eight stereoisomers of a primary subject of this report (BRD50837). c) Dose-response curve of BRD50837 in C3H10T1/2 cell differentiation assay and structure of BRD50837 highlighting positions of building block attachment. All values shown are generated from three independent experiments run in duplicate (values are calculated average ± SD).

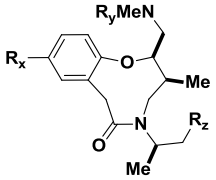
Table 3.1. EC_{50} in C3H10T1/2 cells and PBS solubility of analogs of BRD50837



A



B



C

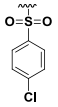
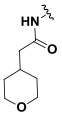
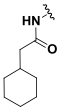
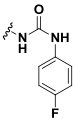
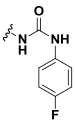
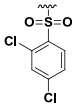
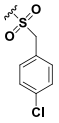
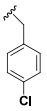
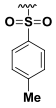
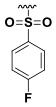
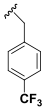
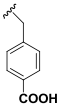
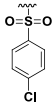
compound	scaffold	R_x	R_y	R_z	EC_{50} (μM)[b]	solubility (μM)[a]
7 (BRD50837)	A				0.09	64.3
8	"		"	"	0.22	-
9	"		"	"	0.29	-
10	"		"	"	0.91	-
11	"		"	"	2.03	-
12	"		"	"	0.08	1.1
19	"		"	"	-	-
20	"		"	"	0.03	5.5
21	"		"	"	0.08	3.9
25	"		"	"	0.18	-
26	"		"	"	0.22	0.7
18	"		"		4.03	-

Table 3.1. (Continued)

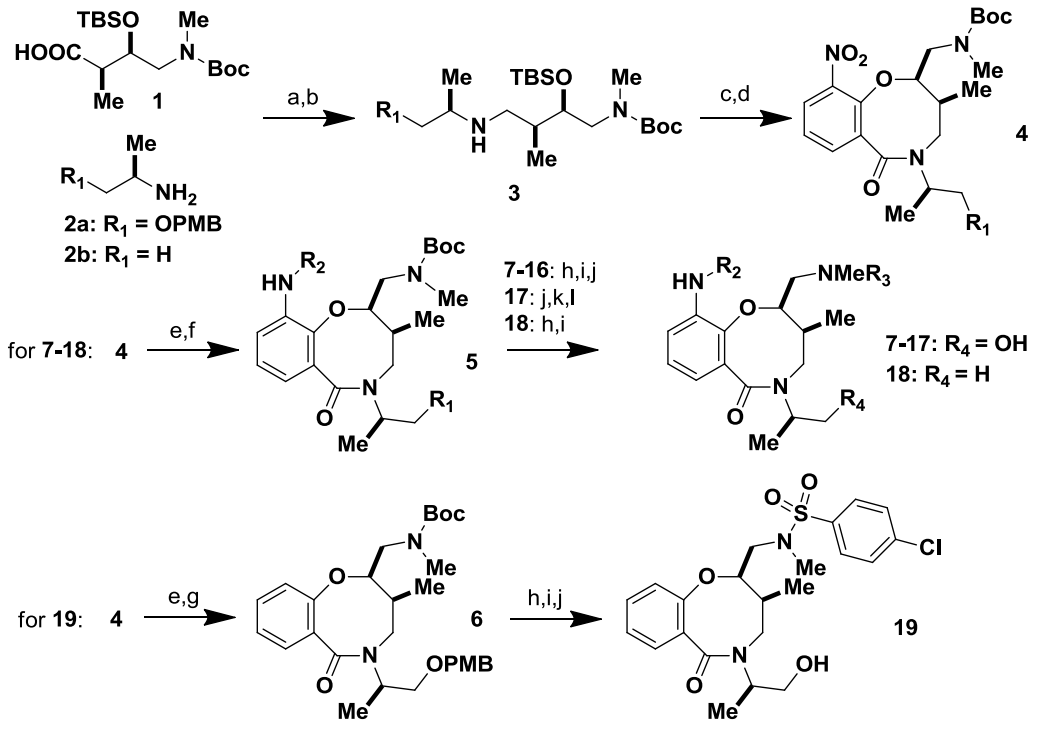
compound	scaffold	R_x	R_y	R_z	EC_{50} (μM)[b]	solubility (μM)[a]
13	"	"			0.95	-
14	"	"		"	1.45	-
15 (BRD9526)	"	"		"	0.06	57.4
16	"	"		"	0.76	-
17	"	"		"	3.37	-
22	"	"		"	0.58	-
23	"	"		"	0.45	-
24	"	"		"	2.86	-
27	"	"		"	2.32	-
28	"	"		"	-	-
29	B	"			-	-
30	C	"	"	"	-	-

a) solubility was measured for compounds that were considered for subsequent experimentation based on their EC_{50}; b) dose-response curve did not pass EC_{50}, so an EC_{50} value was not calculated.

Building block-based structure–activity relationships

To elucidate additional building block-based structure–activity relationships (SAR), I synthesized novel analogs, varying the attachments on the aniline and the extra-annular amine, as well as removing the extra-annular alcohol. BRD50837 (**7**) and additional novel analogs **8-19** were synthesized using an abbreviated synthetic pathway relative to the previously reported solid-phase synthesis (Scheme 3.1, Suppl. Scheme S3.1). For the synthesis of **7-16**, intermediate **4a** was synthesized from **1** and **2a** as previously reported.[2,3] Subsequently, the nitrobenzene was reduced to an aniline that

Scheme 3.1. Synthesis of analogs[a]



[a]Reagents and conditions: (a) PyBOP, DIPEA, DCM, rt, 52-83%. (b) borane dimethyl sulfide complex, THF, 65°C, 88-89%. (c) 2-fluoro-3-nitrobenzoic acid chloride, NEt_3, DCM, rt, 92-100%. (d) CsF, DMF, 85°C, 100%. (e) 10% Pd/C, H_2, EtOH, 35°C, 100%. (f) R_2Cl, 2,6-lutidine, DCM, rt. (g) $NaNO_2$, $NaHSO_3$, EtOH/AcOH, rt, 48%. (h) TBSOTf, 2,6-lutidine, DCM, rt, then HF/pyridine, THF, rt. (i) R_3Cl, 2,6-lutidine, DCM, 0°C, 68% over 3 steps (**18**). (j) DDQ, pH 7 buffer/DCM, rt, 6-59% over 4 steps (**7-16**), 49% over 4 steps (**19**). (k) TFA, DCM, rt. (l) 4-chlorobenzaldehyde, sodium triacetoxy-borohydride, DMF/2% AcOH, rt, 10% over 4 steps (**17**).

was acylated with acyl chlorides to yield anilides **5**. After deprotection of the extra-annular nitrogen with HF/pyridine and addition of a sulfonyl chloride, the PMB group was removed with DDQ, resulting in the final compounds.

To synthesize analog **18**, **2b** was used instead of **2a** as a starting material and the final PMB deprotection step was omitted. Analog **17** was synthesized by preparing intermediate **5a** as before, removing the PMB group with DDQ, deprotecting the Boc group with trifluoroacetic acid (TFA) and forming the tertiary amine **17** by reductive amination with para-chlorobenzaldehyde. Compound **19** was prepared by treating intermediate **4a** with sodium nitrate and sodium bisulfate, which resulted in the deaminated product **6**. Subsequent deprotection of the Boc group with HF/pyridine, addition of the sulfonyl chloride and removal of the PMB group with DDQ yielded **19**. All compounds were purified by column chromatography (30 min, 0-100% ethyl acetate in hexanes) and if necessary by HPLC purification (Synthetic Procedures).

Both new and previously synthesized analogs were tested in C3H10T1/2 cells using Shh-conditioned medium-induced differentiation as readout of Hh signaling (Table 3.1, Suppl. Table S3.1). Viability for all compounds was tested in this system as well, using CellTiter-Glo as a measure of ATP levels (Suppl. Table S3.1). Changes on the aniline moiety of the compound were tolerated, but activity was optimal for saturated ring systems (BRD50837, **12**, **20**). The cyclopropyl derivative (BRD50837) proved to have better solubility in PBS than the cyclohexyl derivative (**12**, **20**), making it the more favorable candidate. Complete removal of the aniline moiety (**19**) resulted in a loss of activity. Ureas instead of amides also showed activity, but were less soluble (**21**, **25**, **26**).

Changing the sulfonamide building block from para-chlorobenzene sulfonyl chloride to ortho- or meta- chlorobenzene sulfonyl chloride (**13**, **14**) resulted in reduced activity. An additional chlorine in the ortho position of the sulfonamide building block (BRD9526 (**15**)) did not impact the activity, suggesting that the chlorine in the para-position is interacting with a putative cellular target, while that in the ortho-position is not. This conclusion was reinforced by the previous observation that the compound lacking the chlorine entirely had reduced activity (**24**). When testing the previously synthesized compounds, it was found that the chlorine derivative is more active as compared to the fluorine and methyl derivatives (**22**, **23**). Having an additional methylene in the sulfonyl chloride also resulted in a decrease of activity (**16**). Preparing the tertiary amine (**17**) rather than the sulfonamide resulted in a decrease of activity, demonstrating a possible electronic or spatial requirement for the sulfonamide connector. This notion was also reflected by additional tertiary amines tested in previous SAR studies (**27**, **28**).

Removing the extra-annular alcohol (**18**) resulted in a loss of activity, showing that the alcohol, which originally was used as a point of attachment in solid-phase synthesis, is necessary.

The original DOS pathway also yielded compounds having eight-membered rings where the aniline moiety is in the para instead of the ortho position, and nine-membered rings where the aniline moiety is also in the para position.[2] BRD50837s analogs having these structural elements were not active (**29**, **30**).

I prioritized BRD50837 and BRD9526 (**15**) for further experimentation, as both displayed good EC_{50} values (Figure 3.3a) and similar PBS solubility (Table 3.1). Neither

showed significant toxicity based on measuring ATP levels as a surrogate for growth or viability (Figure 3.3b). Reduction of *Gli1* expression in C3H10T1/2 cells by 1 μM treatments of these two compounds was confirmed (Suppl. Figure S3.2).

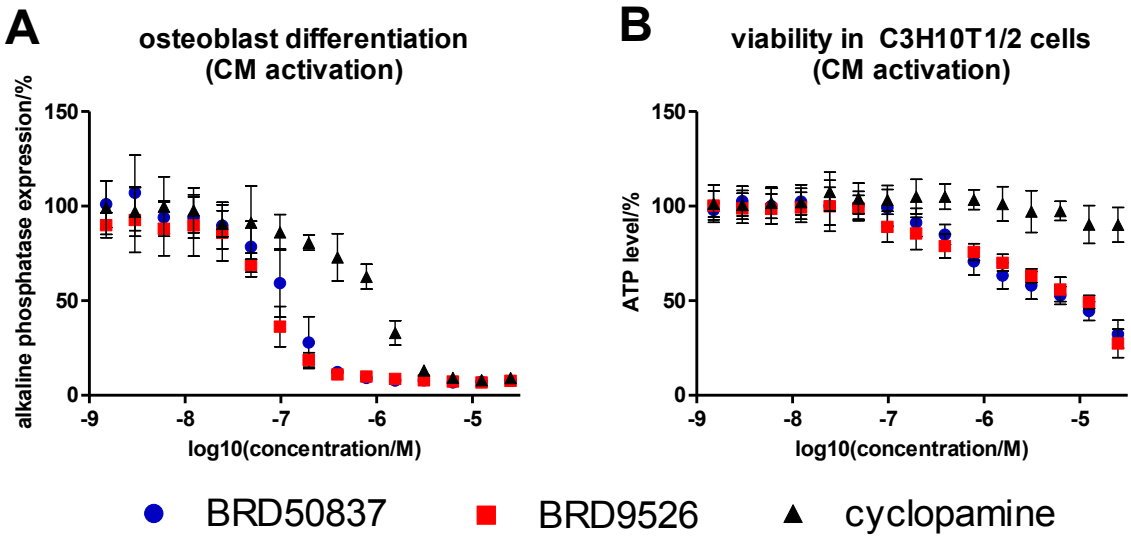


Figure 3.3. All values are shown and generated from three independent experiments run in duplicate (values are calculated average ± SD). (a) Inhibition of Shh-conditioned medium (CM) induced differentiation of C3H10T1/2 cells by BRD50837, BRD9526, and cyclopamine after 48 h. (b) Viability of CM-induced C3H10T1/2 cells in response to 48 h treatment by BRD50837, BRD9526, and cyclopamine.

Conclusion

In conclusion, the Hh pathway inhibitor screen identified a group of compounds with striking correlation between stereochemistry and activity in Hh pathway assays. The stereochemistry-based SAR was confirmed, and further building block-based SAR was conducted. Based on their activity and solubility, two compounds, BRD50837 and BRD9526, were prioritized for further mechanism of action studies.

Experimental Section

Materials and Methods

Screening, Cell Lines and Reagents: as described in Chapter 2.

Organic Synthesis & Compound Characterization: All chemicals and solvents were purchased from Sigma-Aldrich (St Louis, MO, USA), Maybridge (Cornwall, UK), ArkPharm (Libertyville, IL, USA), AstaTech (Bristol, PA, USA), and Aurora (San Diego, CA, USA). Chemicals were used without further purification. Small molecules previously synthesized at the Broad Institute of Harvard and MIT used in SAR studies were purified using HPLC purification. Compound quality (purity and identity) was determined prior to each assay by UPLC-MS (Waters, Milford, MA, USA). Purity was measured by UV absorbance at 210 nm and identity was determined on a SQ mass spectrometer by positive electrospray ionization. Mobile phase A consisted of either 0.01% ammonium hydroxide or 0.01% formic acid in water, while mobile phase B consisted of the same additives in acetonitrile. The gradient ran from 5% to 95% mobile phase B over 0.8 min at 0.45 mL/min. An Acquity BEH C18, 1.7 um, 1.0x50 mm column was used with column temperature maintained at 65°C. Compounds were dissolved in DMSO at a nominal concentration of 1 mg/mL, and 0.25 mL of this solution was injected. Reactions were monitored by thin-layer chromatography (TLC) using E. Merck silica gel 60 F254 precoated plates (0.25 mm) or by tandem liquid chromatography/mass spectroscopy (LCMS) using a Waters 2795 separations module and 3100 mass detector. Visualization for TLC was achieved with UV light and aqueous potassium permanganate ($KMnO_4$) staining followed by heating. NMR spectra were measured on a Bruker Ultrashield 300 (300 MHz ^{1}H, 75 MHz ^{13}C) (Billerica, MA, USA) or Varian Unity

INova 500 (500 MHz ^{1}H) (Palo Alto, CA, USA). Proton chemical shifts are reported in ppm (δ) referenced to the NMR solvent.[4] Data are reported as follows: chemical shifts, multiplicity (br = broad, s = singlet, d = doublet, t = triplet, dd = doublet of doublets, m = multiplet; coupling constant(s) in Hz; integration). All NMR data was collected at 25°C. Infrared spectra were obtained on a Thermo Scientific Nicolet Avatar 370 DTGS IR spectrometer (Waltham, MA, USA) and are reported in cm^{-1}. Optical rotation was measured on a Rudolph Research Analytical Autopol IV automatic polarimeter (Hackettstown, NJ, USA). Flash chromatography purifications were performed using a CombiFlash Rf system (Teledyne ISCO, Inc., Lincoln, NE, USA) with pre-packed 40-60 µm Silica Gel (60 Å mesh) silica gel columns. High resolution LCMS (HRMS) was on an Agilent 1290 Infinity separations module and 6230 time-of-flight (TOF) mass detector operating in ESI+ mode. For HPLC purification, compounds were purified by mass-directed purification on a Waters Autopurification system (Milford, MA, USA). Collection was triggered on the $(M+H)^{+}$ and $(M+Na)^{+}$ ions on a ZQ mass spectrometer using positive electrospray ionization. Mobile phase A consisted of 0.2% ammonium hydroxide in water, while mobile phase B consisted of the 0.2% ammonium hydrocide in acetonitrile. An initial hold at 0% mobile phase B for 1.0 min was followed by a gradient from 0% to 100% mobile phase B over 11.0 min at 24 mL/min. A 2.0 mL/min at-column dilution was present using 100% acetonitrile as well as a 2.0 mL/min make-up flow using 90/10/0.1 methanol/water/formic acid. An XBridge OBD Prep C18, 5 um, 19x100 mm column was used at room temperature. Solubility was determined in phosphate buffered saline (PBS) pH 7.4 with 1% DMSO. Each compound was prepared in duplicate at 100 uM in both 100% DMSO and PBS with 1% DMSO. Compounds were

allowed to equilibrate at room temperature with a 250 rpm orbital shake for 24 hours. After equilibration, samples were analyzed by UPLC-MS (Waters, Milford, MA) with compounds detected by SIR detection on a single quadrupole mass spectrometer. The DMSO samples were used to create a two-point calibration curve to which the response in PBS was fit. Synthetic intermediates were characterized by ^{1}H NMR and LCMS. Final compounds were characterized by ^{1}H NMR and high resolution LCMS, and BRD50837 and BRD9526 were also characterized by ^{13}C NMR, IR, and optical rotation.

Assay Treatments and Readouts: as described in Chapter 2. In addition: RNA extraction reagents, cDNA reverse transcription reagents, and SYBR-Green reagent were purchased from Life Technologies/Applied Biosystems (Carlsbad, CA, USA); primers for qPCR were purchased from Qiagen (Valencia, CA, USA). RNA concentrations were measured using a Thermo Scientific Nanodrop 2000 spectrophotometer (Waltham, MA, USA). qPCR readouts were taken on an ABI Prism 7900 HT qPCR (Carlsbad, CA, USA), using standard curve assay settings (384 wells clear plate) with Sequence Detection Systems Software Version 2.3.

Assay Protocols

Shh Light II screen/assay, Shh Light II CellTiter-Glo Viability assay, C3H10T1/2 differentiation assay, C3H10T1/2 CellTiter-Glo Viability assay: as described in Chapter 2.

C3H10T1/2 cell treatment for gene expression: To confirm that the compounds lower *Gli* expression in C3H10T1/2 cells, gene expression experiments were performed

in these cells. The C3H10T1/2 cells were seeded with 200,000 cells/well into 6 well plates in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. The cells were incubated at 37°C/5% CO_2 until confluent (about 36 hours). Afterwards the wells were washed and DMEM supplemented with 0.5% FBS and 1% penicillin/streptomycin was added to the wells, either with 5% Shh conditioned medium (to those wells induced to differentiate), or without (to those used as non-induced control). The compounds/DMSO were added in the respective concentrations to the Shh conditioned medium induced wells and the respective DMSO amount was added to the non-induced wells. The cells were then incubated for 48 hours at 37°C/5% CO_2. The cells were subsequently harvested by washing them once with PBS, scraping each well and collecting them in 1.7 mL Eppendorf tubes. The tubes were centrifuged at 14,000 rpm for 5 min and the supernatant was carefully aspirated. The pellets were either used right away for RNA extraction or frozen at -20°C until needed.

RNA extraction: If the cell pellets were frozen, they were allowed to thaw for 30 min before starting the RNA extraction. The pellets were resuspended in 1 mL TRIZOL reagent and 200 μL chloroform was added to each sample. The samples were shaken, incubated for 5 min at room temperature and afterwards centrifuged for 15 min at 14,000 rpm. The top, clear phase was carefully extracted and added into a clean 1.7 mL Eppendorf tube. 500 μL isopropanol was added to each sample as well as 5 μL glycogen to precipitate the RNA. The samples were shaken and incubated for 15 min at room temperature and subsequently centrifuged for 15 min at 14,000 rpm. The supernatant was carefully aspirated and the pellet was washed with 1 mL of 70% ethanol in millipore water. The samples were again centrifuged at 14,000 rpm for 10 min

and the supernatant was carefully aspirated. The pellets were air-dried for 30 min and 20 μL TE-buffer or EB-buffer was added. The samples were allowed to resuspend overnight at 4°C or for 30 min at room temperature before the RNA concentrations were measured.

cDNA Reverse Transcription: cDNA Reverse Transcription was done using the High Capacity cDNA Reverse Transcription Kit (ABI) according to manufacturer's instructions with 1 μg RNA per sample. After the reverse transcription was done, the samples were either used right away for the qPCR, or frozen at -20°C until being used for the qPCR.

qPCR: The reverse transcribed samples were diluted with RNAse free water to 55 μL/sample. 11 μL of the respective primer (Gapdh-3 and Gli-1), 22 μL of SYBR Green PCR Master Mix, and 11 μL cDNA of the respective sample were mixed and 10 μL of the mix was pipetted into a 384 well qPCR plate (in quadruplicate). The plates were sealed with optical adhesion film and the qPCR was read out. C_T values were calculated using the instrument's software.

Data Analysis

Primary screening, dose response, and viability data (Shh Light II cells), Dose response curves (C310T1/2): as described in Chapter 2.

EC_{50} dose fits: EC_{50} dose response curves were fitted using a smart fit strategy with initial values of S_0 = 100 and S_{inf} = 10 (C3H10T1/2 differentiation assays), or 0 (viability assays) using Condoseo (Genedata Screener 10.0.2). S_{inf} was chosen based

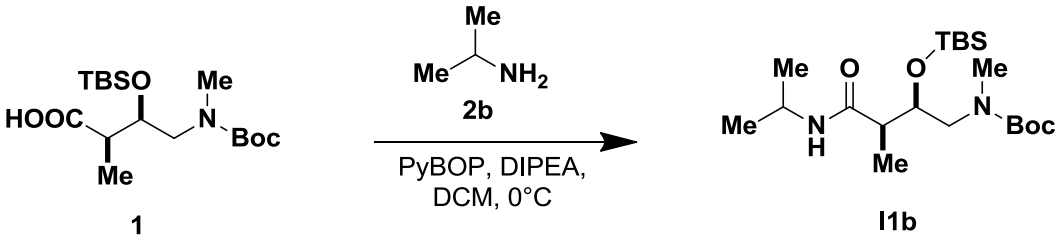
on inhibition of control (12.5 μM cyclopamine for C3H10T1/2 differentiation, and 2.5 μM staurosporine for viability assays).

Gene expression data: For each sample, four technical replicates were measured and C_T values for each replicate were calculated using the instrument's analysis feature. C_T values were subsequently averaged for each sample and the fold expression changes normalized within each experiment to non-induced C3H10T1/2 cells was calculated using the $2^{-\Delta\Delta C}{}_T$ method.[5] Experiments were done three times in triplicate on separate days.

Synthetic Procedures

Intermediate **4a** was synthesized as previously described.[2,3]

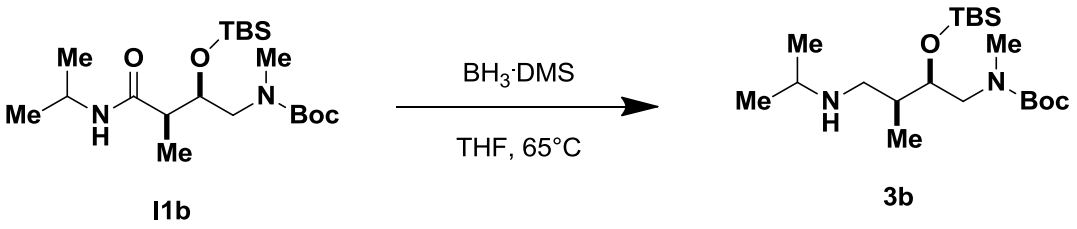
Protocol to Intermediate **4b**:



4-(*tert*-butoxycarbonyl(methyl)amino)-3-(*tert*-butyldimethylsilyloxy)-2-methylbutanoic acid **1** (1.8 g, 1.0 eq) was dissolved in DCM (20 mL, final concentration = 0.2 M) under N_2 and PyBOP (2.64 g, 1.0 eq), and diisopropyl ethylamine (DIPEA) (2.6 mL, 3.0 eq) was added. The mixture was cooled to 0°C before isopropylamine (0.32 g, 1.1 eq) in DCM (4.9 mL) was added. The mixture was allowed to come to room temperature and stirred for 15 h before being quenched with water and extracted with DCM. The combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The

crude product was purified by flash chromatography on silica gel (20-30% EtOAc in hexanes) to give the product **Ib1** with 52% yield.

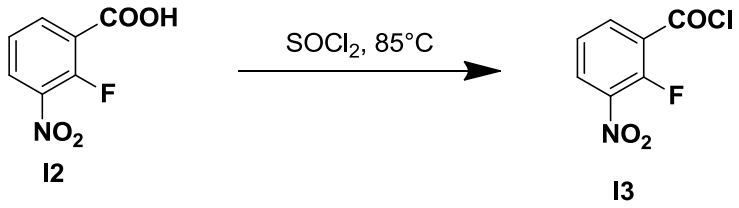
tert-butyl-((2R,3R)-2-((*tert*-butyldimethylsilyl)oxy)-4-(isopropylamino)-3-methyl-4-oxobutyl)(methyl)carbamate **Ib1**: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 6.50 (br dd, *J* = 61.8, 7.1 Hz, 1H), 4.21 – 3.99 (m, 2H), 3.51 (d, *J* = 14.3 Hz , 1H), 3.10 – 2.95 (m, 1H), 2.91 (s, 3H), 2.52 – 2.34 (m, 1H), 1.48 (s, 9H), 1.23 – 1.07 (m, 9H), 0.96 (s, 9H), 0.11 (d, *J* = 10.6 Hz, 6H). LRMS-ES$^+$ *m/z*: 403.35 (M^+).



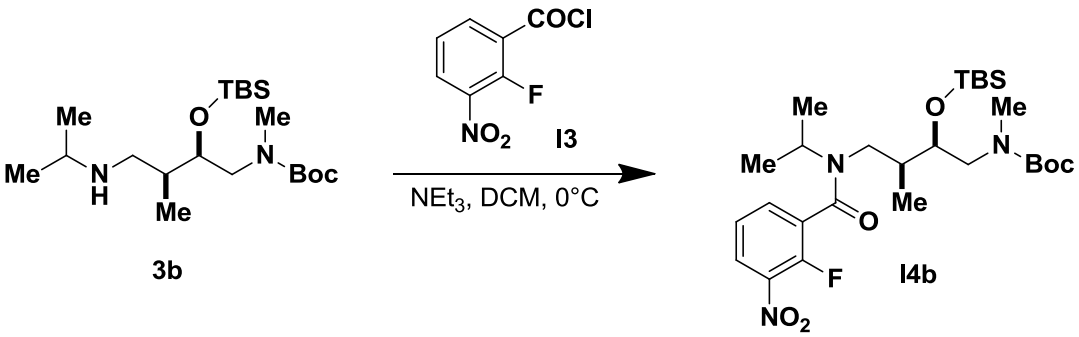
I1b **3b**

tert-butyl-((2R,3R)-2-((*tert*-butyldimethylsilyl)oxy)-4-(isopropylamino)-3-methyl-4-oxobutyl)(methyl)carbamate **I1b** (1.05 g, 1.0 eq) was dissolved in THF (26.1 mL, final concentration 0.1 M) under N_2 and borane dimethylsulfide complex (BH_3·DMS) (1.24 mL, 5.0 eq) was added dropwise. The mixture was heated to 65°C. After 5 h, it was cooled to room temperature and excess hydride was quenched by carefully adding MeOH. The mixture was concentrated under reduced pressure and, to remove excess $B(OMe)_3$, co-evaporated with MeOH three times. The resulting, clear oil was re-dissolved in a 2:3 ratio of MeOH (15.6 mL) and 10% aqueous potassium sodium tartrate (23.3 mL, final concentration 0.067 M) and heated at reflux for 12 h. The product was concentrated under reduced pressure and the aqueous layer extracted three times with EtOAc. The combined organic layers were washed once with brine, dried over $MgSO_4$, filtered, and concentrated to give amine **3b** with 88% yield.

tert-butyl-((2R,3S)-2-((*tert*-butyldimethylsilyl)oxy)-4-(isopropylamino)-3-methylbutyl)(methyl)carbamate **3b**: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 3.98 – 3.83 (m, 1H), 3.47 – 3.28 (m, 1H), 3.02 – 2.90 (m, 1H), 2.86 (s, 3H), 2.77 – 2.65 (m, 1H), 2.65 – 2.50 (m, 1H), 2.42 (dd, *J* = 11.3, 7.4 Hz, 1H), 1.75 – 1.58 (m, 1H), 1.43 (s, 9H), 1.02 (d, *J* = 4.9 Hz, 6H), 0.96 – 0.73 (m, 12H), 0.02 (d, *J* = 5.1 Hz, 6H). LRMS-ES$^+$ *m/z*: 389.40 (M^+).



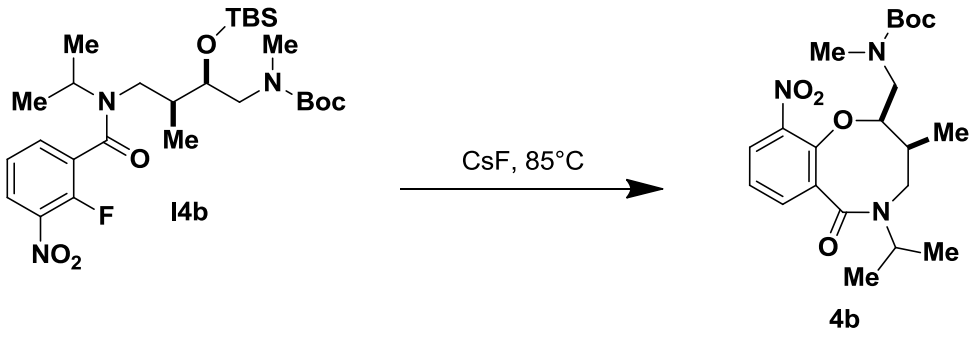
2-fluoro-3-benzoic acid **I2** (7.23 g, 1.0 eq) was dissolved in thionyl chloride ($SOCl_2$) (17.1 mL, 6.0 eq) and heated under reflux under N_2 to 85°C for 2 h. The mixture was then concentrated and the resulting product **I3** immediately used for the next step.



tert-butyl-((2R,3S)-2-((*tert*-butyldimethylsilyl)oxy)-4-(isopropylamino)-3-methylbutyl)(methyl)carbamate **3b** (0.89 g, 1.0 eq) was dissolved in DCM (11.5 mL, 0.2 M) under N_2 and 2-fluoro-3-benzoic acid chloride **I3** (1.17 g, 2.5 eq) was added at 0°C . Triethyl amine (1.6 mL, 5.0 eq) was added and the reaction was brought to room

temperature. The reaction was stirred until complete consumption of **3b** was observed (1-2 h) and subsequently quenched with saturated NH_4Cl solution. The resulting mixture was extracted with DCM and the combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. Flash chromatography on silica gel (10-30 % EtOAc in hexanes) produced the product **I4b** with 92% yield.

tert-butyl-((2R,3S)-2-((*tert*-butyldimethylsilyl)oxy)-4-(2-fluoro-N-isopropyl-3-nitrobenzamido)-3-methylbutyl)(methyl)carbamate **I4b**: ^{1}H NMR (300 MHz, $CDCl_3$, room temperatureδ 8.13 – 7.98 (m, 1H), 7.70 – 7.53 (m, 1H), 7.41 – 7.29 (m, 1H), 4.04 – 3.91 (m, 1H), 3.83 – 3.41 (m, 3H), 3.41 – 3.00 (m, 3H), 2.95 – 2.68 (m, 3H), 2.49 – 2.27 (m, 1H), 2.27 – 2.08 (m, 1H), 1.91 – 1.78 (m, 1H), 1.69 – 1.54 (m, 1H), 1.50 – 1.31 (m, 9H),1.15 – 1.04 (m, 2H), 1.04 – 0.94 (m, 2H), 0.94 – 0.87 (m, 6H), 0.82 – 0.73 (m, 1H) 0.64 (s, 3H), 0.10 (d, *J* = 10.9 Hz, 4H), -0.03 (s, 1H), -0.15 (s, 1H). LRMS-ES$^+$ *m/z*: 578.33 $(M+Na)^+$.

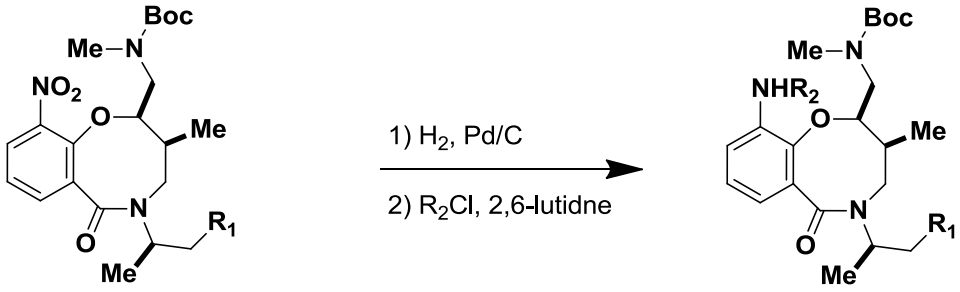


tert-butyl-((2R,3S)-2-((*tert*-butyldimethylsilyl)oxy)-4-(2-fluoro-N-isopropyl-3-nitrobenzamido)-3-methylbutyl)(methyl)carbamate **I4b** (1.17 g, 1.0 eq) was dissolved in DMF (21.1 mL, 0.1 M) at room temperature under N_2 and cesium fluoride (0.96 g, 3 eq) was added. The reaction was stirred and heated to 85°C. After complete consumption

of starting material was observed (~5h), the reaction was cooled and concentrated under reduced pressure. The resulting oil was taken up in EtOAc, washed with saturated NH_4Cl solution, dried with $MgSO_4$, filtered, and concentrated to yield **4b**. The product was used without further purification.

tert-butyl-(((2R,3S)-5-isopropyl-3-methyl-10-nitro-6-oxo-3,4,5,6-tetrahydro-2H-benzo[b][1,5]oxazocin-2-yl)methyl)(methyl)carbamate **4b**: ^1H NMR (300 MHz, $CDCl_3$, room temperature) δ 7.76 – 7.57 (m, 2H), 7.03 (t, *J* = 7.9 Hz, 1H), 4.95 – 4.67 (m, 2H), 3.90 – 3.57 (m, 1H), 3.40 – 3.01 (m, 3H), 2.88 (s, 3H), 2.20 – 1.97 (m, 1H), 1.44 (s, 9H), 1.36 – 1.19 (m, 6H), 1.09 – 0.96 (m, 3H). LRMS-ES^+ *m/z*: 444.31 $(M+Na)^+$.

General Protocol to Intermediate **5**:



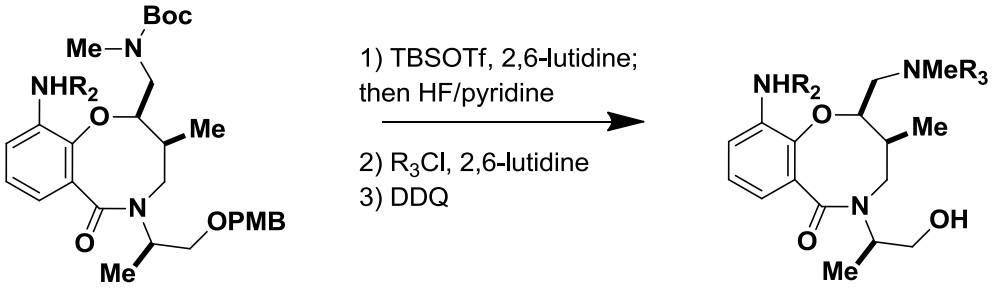
4a: R_1 = OPMB
4b: R_1 = H

5

4 (1.0 eq) in EtOH (0.05 M) was stirred with palladium (10% on activated carbon, 0.10 eq) at 35°C under H_2. After the starting material was completely consumed (as monitored by LC-MS, 1-2 h), the mixture was cooled, filtered through celite, and concentrated. The crude material was dissolved in DCM (final concentration 0.02 M) and the desired acyl chloride (1.1 eq, see Suppl. Scheme S3.1 for acyl chlorides used) was added. The mixture was cooled to 0°C and 2,6-lutidine (5.0 eq) was added. The reaction was stirred overnight and subsequently quenched with saturated NH_4Cl

solution. The resulting mixture was extracted with DCM and the combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The product was immediately used for the next step without further purification.

General Protocol from **5a** to **7**-**16**:



5a **7-16**

5a (1.0 eq) was dissolved in DCM (0.1 M) and 2,6-lutidine (4.0 eq) and TBSOTf (3.0 eq) were added at room temperature. The reaction was stirred for 2h and subsequently quenched with saturated NH_4Cl solution. The mixture was extracted with EtOAc and the combined organic extracts were dried over $MgSO_4$, filtered, and concentrated to produce the crude silyl carbamate which was dissolved again in THF (0.1 M). HF/pyridine (70%, 1.0 eq) was added and the mixture was stirred for 40 min at room temperature. The reaction was then quenched with saturated NH_4Cl solution and extracted with EtOAc. The combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The crude material was dissolved in DCM (0.03 M) and 2,6-lutidine (6.0 eq) was added and the mixture was cooled to 0°C. The desired sulfonyl chloride (3.0 eq, see Suppl. Scheme S3.1 for sulfonyl chlorides used) was added in minimal DCM. The reaction was stirred at 0°C overnight and subsequently quenched with water. The mixture was extracted with DCM, dried over $MgSO_4$, filtered, and concentrated. The

product was filtered through a short silica plug (0-100 % EtOAc in Hexanes) to remove polar by-products. The product was then dissolved in a 4:1 DCM and pH 7 buffer solution (final concentration ranging from 0.03 M to 0.04 M). The mixture was cooled to 0°C and DDQ (1.5 eq) was added. The reaction was stirred at 0°C for another 10 mins and then brought to room temperature and stirred overnight. The reaction was quenched with water and extracted with DCM. The combined organic extracts were washed with saturated $NaHCO_3$ solution and subsequently, activated carbon was added. The mixture was filtered through celite and the filtrand was washed with hot DCM several times. The filtrate was concentrated and the product was purified by flash chromatography on silica gel (0-100% EtOAc in hexanes). If the compound was not pure by LCMS, it was further purified as mentioned.

BRD50837, (**7**)

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 1 g of the resulting product was reacted with 0.19 mL **AC1** and 1.1 mL 2,6-lutidine in 95 mL DCM. 0.7 g of the product was deprotected using 0.81 mL TBSOTf, 0.55 mL 2,6-lutidine in 11.8 mL DCM, and 0.11 mL HF/pyridine in 11.8 mL THF. 0.15 g of the product was reacted with 0.19 g **SC1** and 0.21 mL 2,6-lutidine in 12.1 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M) to give **7** in 24% overall yield.

BRD50837 (**7**): $[\alpha]_D^{20}$ 57.8 (*c* 1.0, $CHCl_3$). IR (cm^{-1}) 3365 (br, w), 2925 (w), 1680 (m), 1610 (s), 1578 (m), 1529 (s), 1434 (s), 1337 (s), 1159 (s), 1093 (m), 975 (s), 767 (s),

752 (s), 609 (m), 558 (m). ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.81 (s, 1H), 8.38 (d, *J* = 7.8 Hz, 1H), 7.70 (d, *J* = 8.4 Hz, 2H), 7.48 (d, *J* = 8.5 Hz, 2H), 7.10 (d, *J* = 7.9 Hz, 1H), 6.92 (t, *J* = 8.0 Hz, 1H), 5.07 – 4.85 (m, 2H), 4.17 – 3.97 (m, 1H), 3.87 (dd, *J* = 11.7, 3.5 Hz, 1H), 3.56 – 3.40 (m, 1H), 3.36 – 3.11 (m, 3H), 2.74 – 2.55 (m, 4H), 2.06 – 1.92 (m, 1H), 1.92 – 1.80 (m, 1H), 1.18 (d, *J* = 6.9 Hz, 3H), 1.08 – 0.94 (m, 5H), 0.84 – 0.74 (m, 2H). ^{13}C NMR (75 MHz, $CDCl_3$, room temperature) δ 173.1, 172.1, 144.3, 139.4, 136.7, 129.6, 128.6, 128.5, 126.0, 122.6, 121.0, 120.7, 72.8, 66.5, 53.0, 52.4, 47.6, 35.4, 35.1, 15.4, 14.2, 11.2, 8.2, 8.1. HRMS (ESI) calcd for $C_{26}H_{33}ClN_3O_6S$ $[M+H]^+$: 550.1779. Found: 550.1783.

8

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 0.1 g of the resulting product was reacted with 0.015 mL **AC2** and 0.11 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.13 mL TBSOTf, 0.09 mL 2,6-lutidine in 1.9 mL DCM, and 0.017 mL HF/pyridine in 1.9 mL THF. The product was reacted with 0.12 g **SC1** and 0.13 mL 2,6-lutidine in 7.7 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M) to give **8** in 36% overall yield.

8: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.64 (s, 1H), 8.41 (d, *J* = 6.9 Hz, 1H), 7.71 (d, *J* = 8.6 Hz, 2H), 7.49 (d, *J* = 8.6 Hz, 2H), 7.13 (d, *J* = 8.0 Hz, 1H), 6.95 (t, *J* = 8.0 Hz, 1H), 5.05 – 4.85 (m, 2H), 4.04 (dd, *J* = 14.2, 10.1 Hz, 1H), 3.88 (dd, *J* = 11.5, 2.0 Hz, 1H), 3.55 – 3.40 (m, 1H), 3.36 – 3.23 (m, 1H), 3.03 (d, *J* = 8.6 Hz, 1H), 2.70 (s, 3H), 2.62 (d, *J* = 14.1 Hz, 1H), 2.23 (s, 3H), 2.06 – 1.90 (m, 1H), 1.78 (s, 1H), 1.19 (d, *J*

= 7.0 Hz, 3H), 0.99 (d, *J* = 6.7 Hz, 3H). HRMS (ESI) calcd for $C_{24}H_{31}ClN_3O_6S$ $[M+H]^+$: 524.1622. Found: 524.1624.

9

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 0.1 g of the resulting product was reacted with 0.022 mL **AC3** and 0.11 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.13 mL TBSOTf, 0.09 mL 2,6-lutidine in 1.9 mL DCM, and 0.017 mL HF/pyridine in 1.9 mL THF. The product was reacted with 0.12 g **SC1** and 0.13 mL 2,6-lutidine in 7.5 mL DCM and the product was deprotected using 0.06 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.04 M) to give **9** in 36% overall yield.

9: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.87 (s, 1H), 8.46 (d, *J* = 7.9 Hz, 1H), 8.02 (d, *J* = 6.8 Hz, 2H), 7.63 (d, *J* = 8.5 Hz, 2H), 7.57 – 7.34 (m, 5H), 7.20 (d, *J* = 8.0 Hz, 1H), 7.02 (t, *J* = 8.0 Hz, 1H), 5.09 – 4.86 (m, 2H), 3.95 – 3.79 (m, 2H), 3.55 – 3.40 (m, 1H), 3.34 – 3.22 (m, 1H), 3.17 (d, *J* = 8.5 Hz, 1H), 2.75 (d, *J* = 14.1 Hz, 1H), 2.69 (s, 3H), 2.09 – 1.93 (m, 1H), 1.89 (s, 1H) , 1.18 (d, *J* = 6.9 Hz, 3H), 0.96 (d, *J* = 6.6 Hz, 3H). HRMS (ESI) calcd for $C_{29}H_{33}ClN_3O_6S$ $[M+H]^+$: 586.1779. Found: 586.1784.

10

Following the general protocol, 3.58 g of **4a** was reacted with 0.68 g palladium (10% on activated carbon) in 128 mL EtOH. 0.1 g of the resulting product was reacted with 0.03 g **AC4** and 0.11 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.15 mL TBSOTf, 0.10 mL 2,6-lutidine in 2.2 mL DCM, and 0.020 mL HF/pyridine in 2.2

mL THF. The product was reacted with 0.13 g **SC1** and 0.15 mL 2,6-lutidine in 8.8 mL DCM and the product was deprotected using 0.07 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.04 M) to give **10** in 23% overall yield.

10: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature, 2 diastereomers) δ 9.03 (d, *J* = 4.3 Hz, 1H), 8.42 (t, *J* = 7.5 Hz, 1H), 7.67 (dd, *J* = 8.5, 2.8 Hz, 2H), 7.44 (d, *J* = 8.5 Hz, 2H), 7.17 (d, *J* = 7.9 Hz, 1H), 6.95 (t, *J* = 8.0 Hz, 1H), 5.07 – 4.92 (m, 1H), 4.86 – 4.68 (m, 1H), 4.17 – 4.00 (m, 1H), 3.90 (d, *J* = 10.9 Hz, 2H), 3.64 – 3.47 (m, 2H), 3.42 (dd, *J* = 14.1, 4.0 Hz, 1H), 3.31 – 3.18 (m, 2H), 3.12 (d, *J* = 6.0 Hz, 1H), 2.78 (d, *J* = 6.6 Hz, 3H), 2.26 – 2.08 (m, 2H), 1.98 – 1.88 (m, 1H), 1.86 – 1.78 (s, 1H), 1.63 – 1.50 (m, 4H), 1.29 – 1.14 (m, 3H), 1.02 – 0.94 (m, 3H). HRMS (ESI) calcd for $C_{28}H_{37}ClN_3O_7S$ $[M+H]^+$: 594.2041. Found: 594.2048.

11

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 0.1 g of the resulting product was reacted with 0.03 mL **AC5** and 0.10 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.14 mL TBSOTf, 0.09 mL 2,6-lutidine in 2.0 mL DCM, and 0.018 mL HF/pyridine in 2.0 mL THF. The product was reacted with 0.13 g **SC1** and 0.14 mL 2,6-lutidine in 8.0 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M) to give **11** in 24% overall yield.

11: ^{1}H NMR (500 MHz, DMSO-d_6, room temperature) δ 8.73 (s, 1H), 8.10 (d, *J* = 7.8 Hz, 1H), 7.84 (d, *J* = 8.6 Hz, 2H), 7.70 (d, *J* = 8.6 Hz, 2H), 7.05 (d, *J* = 7.9 Hz, 1H), 6.90 (t, *J* = 7.9 Hz, 1H), 4.91 (t, *J* = 5.1 Hz, 1H), 4.83 (d, *J* = 8.6 Hz, 1H), 4.68 – 4.58 (m, 1H),

3.80 (t, *J* = 10.5 Hz, 2H), 3.75 – 3.64 (m, 1H), 3.59 – 3.44 (m, 2H), 3.30 – 3.22 (m, 2H), 3.06 (d, *J* = 14.1 Hz, 1H), 2.97 (dd, *J* = 15.0, 12.7 Hz, 1H), 2.69 (s, 3H), 2.33 (d, *J* = 6.8 Hz, 2H), 2.12 – 2.03 (m, 1H), 2.02 – 1.95 (m, 1H), 1.57 (dd, *J* = 27.2, 12.6 Hz, 2H), 1.29 – 1.17 (m, 3H), 1.10 (d, *J* = 6.9 Hz, 3H), 0.89 (d, *J* = 6.6 Hz, 3H). HRMS (ESI) calcd for $C_{29}H_{39}ClN_3O_7S$ $[M+H]^+$: 608.2197. Found: 608.2204.

12

Following the general protocol, 0.51 g of **4a** was reacted with 0.10 g palladium (10% on activated carbon) in 18.4 mL EtOH. 0.12 g of the resulting product was reacted with 0.04 mL **AC6** and 0.13 mL 2,6-lutidine in 11.4 mL DCM. The product was deprotected using 0.16 mL TBSOTf, 0.11 mL 2,6-lutidine in 2.3 mL DCM, and 0.021 mL HF/pyridine in 2.3 mL THF. The product was reacted with 0.15 g **SC1** and 0.17 mL 2,6-lutidine in 9.4 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M). The final product was additionally purified by preparative thin layer chromatography using E. Merck silica gel 60 F254 precoated plates (2 mm) and 100% EtOAc to give **12** in 6% overall yield.

12: ^{1}H NMR (500 MHz, DMSO-d_6, room temperature) δ 8.68 (s, 1H), 8.08 (d, *J* = 7.7 Hz, 1H), 7.83 (d, *J* = 8.5 Hz, 2H), 7.70 (d, *J* = 8.5 Hz, 2H), 7.04 (d, *J* = 7.9 Hz, 1H), 6.90 (t, *J* = 7.9 Hz, 1H), 4.91 (t, *J* = 5.0 Hz, 1H), 4.83 (d, *J* = 8.3 Hz, 1H), 4.67 – 4.59 (m, 1H), 3.65 (dd, *J* = 14.1, 9.2 Hz, 1H), 3.56 – 3.47 (m, 2H), 3.42 – 3.35 (m, 1H), 3.05 (d, *J* = 14.0 Hz, 1H), 3.00 – 2.92 (m, 1H), 2.69 (s, 3H), 2.29 – 2.20 (m, 2H), 2.12 – 2.04 (m, 1H), 1.75 – 1.68 (m, 2H), 1.66 – 1.59 (m, 3H), 1.24 – 1.12 (m, 4H), 1.09 (d, *J* = 6.8 Hz,

3H), 0.99 – 0.92 (m, 2H), 0.88 (d, *J* = 6.4 Hz, 3H). HRMS (ESI) calcd for $C_{30}H_{41}ClN_3O_6S$ $[M+H]^+$: 606.2405. Found: 606.2410.

13

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 1 g of the resulting product was reacted with 0.19 mL **AC1** and 1.1 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.81 mL TBSOTf, 0.55 mL 2,6-lutidine in 11.8 mL DCM, and 0.11 mL HF/pyridine in 11.8 mL THF. 0.15 g of the product was reacted with 0.13 mL **SC2** and 0.21 mL 2,6-lutidine in 12.1 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M) to give **13** in 19% overall yield.

13: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.81 (s, 1H), 8.39 (d, *J* = 7.9 Hz, 1H), 7.76 (s, 1H), 7.65 (d, *J* = 7.7 Hz, 1H), 7.55 (d, *J* = 8.0 Hz, 1H), 7.45 (t, *J* = 7.9 Hz, 1H), 7.10 (d, *J* = 8.0 Hz, 1H), 6.93 (t, *J* = 8.0 Hz, 1H), 5.05 – 4.85 (m, 2H), 4.16 – 4.00 (m, 1H), 3.88 (dd, *J* = 11.6, 3.7 Hz, 1H), 3.54 – 3.40 (m, 1H), 3.36 – 3.18 (m, 2H), 3.11 (d, *J* = 8.7 Hz, 1H), 2.74 – 2.61 (m, 4H), 2.05 – 1.93 (m, 1H), 1.93 – 1.77 (m, 1H), 1.19 (d, *J* = 6.9 Hz, 3H), 1.08 – 0.93 (m, 5H), 0.85 – 0.72 (m, 2H). HRMS (ESI) calcd for $C_{26}H_{33}ClN_3O_6S$ $[M+H]^+$: 550.1779. Found: 550.1782.

14

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 1 g of the resulting product was reacted with 0.19

mL **AC1** and 1.1 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.81 mL TBSOTf, 0.55 mL 2,6-lutidine in 11.8 mL DCM, and 0.11 mL HF/pyridine in 11.8 mL THF. 0.15 g of the product was reacted with 0.12 mL **SC3** and 0.21 mL 2,6-lutidine in 12.1 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M) to give **14** in 35% overall yield.

14: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.83 (s, 1H), 8.34 (d, *J* = 7.9 Hz, 1H), 7.99 (d, *J* = 7.9 Hz, 1H), 7.54 – 7.40 (m, 2H), 7.34 (t, *J* = 7.3 Hz, 1H), 7.06 (d, *J* = 8.0 Hz, 1H), 6.88 (t, *J* = 8.0 Hz, 1H), 5.10 (d, *J* = 9.9 Hz, 1H), 5.03 – 4.85 (m, 1H), 4.51 – 4.33 (m, 1H), 3.87 (dd, *J* = 11.6, 3.5 Hz, 1H), 3.58 – 3.10 (m, 1H), 3.40 – 3.12 (m, 3H), 2.96 (d, *J* = 14.7 Hz, 1H), 2.66 (s, 3H), 2.15 – 1.93 (m, 1H), 1.93 – 1.78 (m, 1H), 1.18 (d, *J* = 6.9 Hz, 3H), 1.06 – 0.91 (m, 5H), 0.84 – 0.67 (m, 2H). HRMS (ESI) calcd for $C_{26}H_{33}ClN_3O_6S$ $[M+H]^+$: 550.1779. Found: 550.1781.

BRD9526 (**15**)

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 1 g of the resulting product was reacted with 0.19 mL **AC1** and 1.1 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.81 mL TBSOTf, 0.55 mL 2,6-lutidine in 11.8 mL DCM, and 0.11 mL HF/pyridine in 11.8 mL THF. 0.15 g of the product was reacted with 0.23g **SC4** and 0.21 mL 2,6-lutidine in 12.1 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M) to give **15** in 20% overall yield.

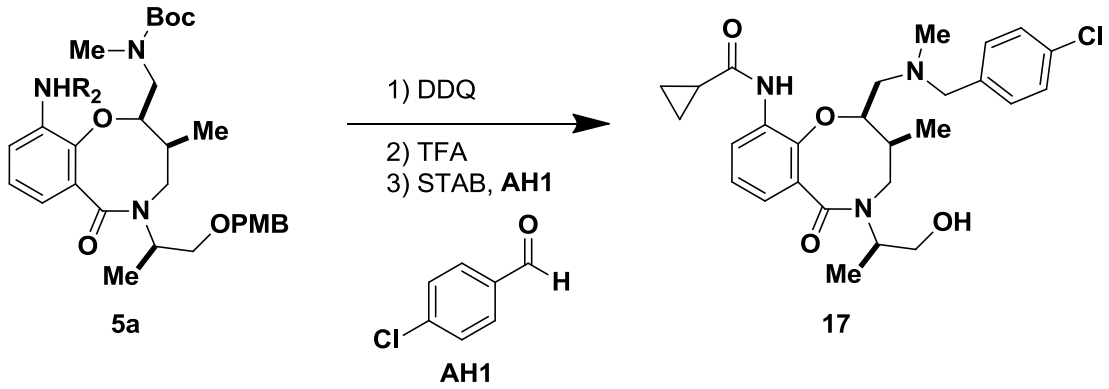
BRD9526 (**15**): $[\alpha]_D^{20}$ 8.2 (*c* 1.0, $CHCl_3$). IR (cm^{-1}) 3366 (br, w), 2929 (w), 1680 (m), 1610 (s), 1576 (m), 1530 (s), 1434 (s), 1334 (s), 1158 (s), 1038 (m), 978 (s), 821 (s), 750 (s), 612 (s) 492 (m). ^1H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.77 (s, 1H), 8.34 (d, *J* = 7.8 Hz, 1H), 7.92 (d, *J* = 8.5 Hz, 1H), 7.50 (s, 1H), 7.32 (d, *J* = 8.5 Hz, 1H), 7.05 (d, *J* = 7.8 Hz, 1H), 6.88 (t, *J* = 7.9 Hz, 1H), 5.11 (d, *J* = 9.9 Hz, 1H), 5.05 – 4.90 (m, 1H), 4.39 (dd, *J* = 14.6, 10.4 Hz, 1H), 3.87 (dd, *J* = 11.0, 3.4 Hz, 1H), 3.56 – 3.35 (m, 2H), 3.28 (d, *J* = 9.1 Hz, 2H), 2.96 (d, *J* = 14.6 Hz, 1H), 2.63 (s, 3H) 2.11 – 1.94 (m, 1H), 1.94 – 1.77 (m, 1H), 1.17 (d, *J* = 6.8 Hz, 3H), 1.07 – 0.91 (m, 5H), 0.84 – 0.70 (m, 2H). ^{13}C NMR (75 MHz, $CDCl_3$, room temperature) δ 173.2, 172.6, 144.1, 139.7, 135.4, 133.0, 132.9, 131.9, 128.5, 127.5, 125.6, 122.6, 121.1, 120.6, 71.9, 66.5, 53.4, 52.0, 47.4, 35.4, 34.0, 15.5, 14.3, 11.3, 8.2, 8.2. HRMS (ESI) calcd for $C_{26}H_{32}Cl_2N_3O_6S$ $[M+H]^+$: 584.1389. Found: 584.1396.

16

Following the general protocol, 1.46 g of **4a** was reacted with 0.28 g palladium (10% on activated carbon) in 52.4 mL EtOH. 1 g of the resulting product was reacted with 0.19 mL **AC1** and 1.1 mL 2,6-lutidine in 9.5 mL DCM. The product was deprotected using 0.81 mL TBSOTf, 0.55 mL 2,6-lutidine in 11.8 mL DCM, and 0.11 mL HF/pyridine in 11.8 mL THF. 0.15 g of the product was reacted with 0.20 g **SC5** and 0.21 mL 2,6-lutidine in 12.1 mL DCM and the product was deprotected using 0.05 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution (final concentration: 0.03 M) to give **16** in 26% overall yield.

16: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.78 (s, 1H), 8.40 (d, *J* = 7.0 Hz, 1H), 7.34 (d, *J* = 8.3 Hz, 2H), 7.26 (d, *J* = 1.1 Hz, 2H), 7.05 (d, *J* = 7.9 Hz, 1H), 6.91 (t, *J* = 8.0 Hz, 1H), 5.05 – 4.86 (m, 2H), 4.16 (s, 2H), 3.93 – 3.73 (m, 2H), 3.56 – 3.38 (m, 1H), 3.35 – 3.14 (m, 3H), 2.68 – 2.51 (m, 4H), 2.07 – 1.88 (m, 1H), 1.88 – 1.77 (m, 1H), 1.16 (d, *J* = 6.9 Hz, 3H), 1.06 – 0.96 (m, 2H), 0.91 – 0.76 (m, 5H). HRMS (ESI) calcd for $C_{27}H_{35}ClN_3O_6S$ $[M+H]^+$: 564.1935. Found: 564.1939.

General Protocol from **5a** to **17**:



5a (1.0 eq) was dissolved in a 4:1 DCM and pH 7 buffer solution (final concentration: 0.04 M). The mixture was cooled to 0°C and DDQ (1.5 eq) was added. The reaction was stirred at 0°C for another 10 mins and then brought to room temperature and stirred for 4 h. The reaction was quenched with water and extracted with DCM. The combined organic extracts were washed with saturated $NaHCO_3$ solution and subsequently, activated carbon was added. The mixture was filtered through celite and the filtrand was washed with hot DCM several times. The product was filtered through a short silica plug (0-100 % EtOAc in Hexanes) to remove polar by-products. The product was then dissolved in DCM (1.0 M) and TFA (0.1 eq to 1.0 eq DCM) was added. The reaction was stirred for 1 h at room temperature. The mixture was concentrated and saturated

$NaHCO_3$ solution was added. The mixture was extracted with DCM, and the combined organic extracts were dried with $MgSO_4$, filtered, and concentrated. The product was dissolved in DMF with 2% acetic acid (0.03 M) and the aldehyde **AH1** (3.0 eq) was added. The mixture was stirred for 45 mins before sodium triacetoxyborohydride (STAB, 3.0 eq) was added. The reaction was stirred overnight and in the morning, another 2.0 eq of **AH1** and STAB were added and the reaction was allowed to stir for another 2 h. The reaction was diluted with EtOAc and $NaHCO_3$ and extracted with EtOAc. The mixture was washed with water and brine and the combined organic extracts were dried with Na_2SO_4, filtered, and concentrated. The product was purified by flash chromatography on silica gel (0-100% EtOAc in hexanes).

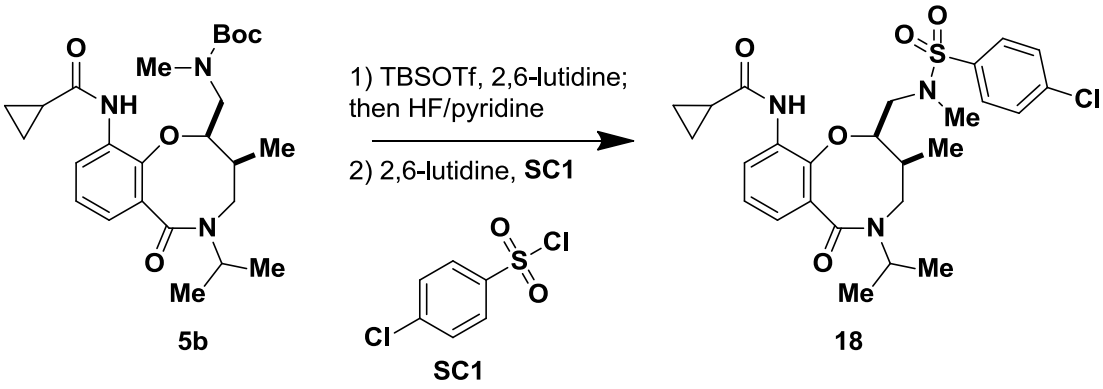
17

Following the general protocol, 0.51 g of **4a** was reacted with 0.10 g palladium (10% on activated carbon) in 18.4 mL EtOH. 0.11 g of the resulting product was reacted with 0.02 mL **AC1** and 0.12 mL 2,6-lutidine in 10.4 mL DCM. The product was reacted with 0.07 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution before 0.06 g of the product was deprotected in a 1:10 dilution of 0.12 mL TFA in 1.15 mL DCM. The product was then reacted with 0.09 g STAB and 0.6 g **AH1**, giving **17** after purification in 10% overall yield (91% purity).

17: ^{1}H NMR (500 MHz, DMSO-d_6, room temperature, 91% purity) δ 9.02 (s, 1H), 7.90 (d, *J* = 7.3 Hz, 1H), 7.33 (d, *J* = 8.3 Hz, 2H), 7.26 (d, *J* = 8.3 Hz, 2H), 7.10 (d, *J* = 7.8 Hz, 1H), 6.85 (t, *J* = 8.0 Hz, 1H), 4.88 (t, *J* = 5.0 Hz, 1H), 4.74 – 4.67 (m, 1H), 4.50 (dd, *J* = 13.2, 6.8 Hz, 1H), 3.60 – 3.55 (m, 1H), 3.55 – 3.50 (m, 2H), 3.50 – 3.45 (m, 1H),

3.31 – 3.26 (m, 1H), 3.02 – 2.90 (m, 2H), 2.62 (dd, *J* = 13.3, 3.2 Hz, 1H), 2.11 (s, 3H), 2.02 – 1.93 (m, 1H), 1.11 (d, *J* = 6.9 Hz, 3H), 0.85 (d, *J* = 6.6 Hz, 3H), 0.82 – 0.65 (m, 5H).HRMS (ESI) calcd for $C_{27}H_{35}ClN_3O_4$ $[M+H]^+$: 500.2316. Found: 500.2317.

Protocol from **5b** to **18**:



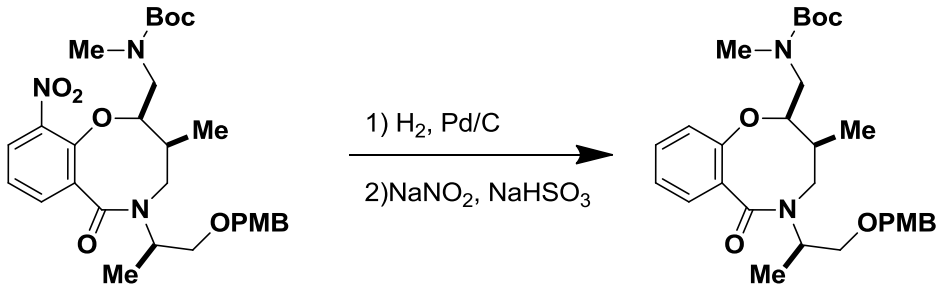
5b (1.0 eq) was dissolved in DCM (0.1 M) and 2,6-lutidine (4.0 eq) and TBSOTf (3.0 eq) were added at room temperature. The reaction was stirred for 2h and subsequently quenched with saturated NH_4Cl solution. The mixture was extracted with EtOAc and the combined organic extracts were dried over $MgSO_4$, filtered, and concentrated to produce the crude silyl carbamate which was dissolved again in THF (0.1 M). HF/pyridine (70%, 1.0 eq) was added and the mixture was stirred for 40 min at room temperature. The reaction was then quenched with saturated NH_4Cl solution and extracted with EtOAc. The combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The crude material was dissolved in DCM (0.03 M) and 2,6-lutidine (6.0 eq) was added and the mixture was cooled to 0°C. The desired sulfonyl chloride (3.0 eq) was added in minimal DCM. The reaction was stirred at 0°C overnight and subsequently quenched with water. The mixture was extracted with DCM, dried over $MgSO_4$, filtered, and concentrated. The compound was purified by HPLC.

18

Following the general protocol, 0.96 g of **4b** was reacted with 0.24 g palladium (10% on activated carbon) 46 mL EtOH. The resulting product was reacted with 0.23 mL **AC1** and 1.3 mL 2,6-lutidine in 114 mL DCM. 0.45 g of the resulting product was deprotected using 0.66 mL TBSOTf, 0.47 mL 2,6-lutidine in 9.8 mL DCM, and 0.10 mL HF/pyridine in 9.8 mL THF. The product was reacted with 0.62 g **SC1** and 0.68 mL 2,6-lutidine in 38.9 mL DCM to give **18** in 68% overall yield.

18: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 8.78 (s, 1H), 8.36 (d, *J* = 7.8 Hz, 1H), 7.69 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.3 Hz, 2H), 7.13 (d, *J* = 7.6 Hz, 1H), 6.90 (t, *J* = 7.9 Hz, 1H), 4.85 – 4.70 (m, 1H), 4.64 (d, *J* = 9.5 Hz, 1H), 4.13 – 3.96 (m, 1H), 3.23 (d, *J* = 8.7 Hz, 2H), 2.72 (s, 3H), 2.65 (d, *J* = 14.2 Hz, 1H), 1.96 – 1.76 (m, 2H), 1.26 – 1.15 (m, 6H), 1.05 – 0.91 (m, 5H), 0.81 – 0.69 (m, 2H). HRMS (ESI) calcd for $C_{26}H_{33}ClN_3O_5S$ $[M+H]^+$: 534.1829. Found: 534.1836.

General Protocol from **4a** to **19**:



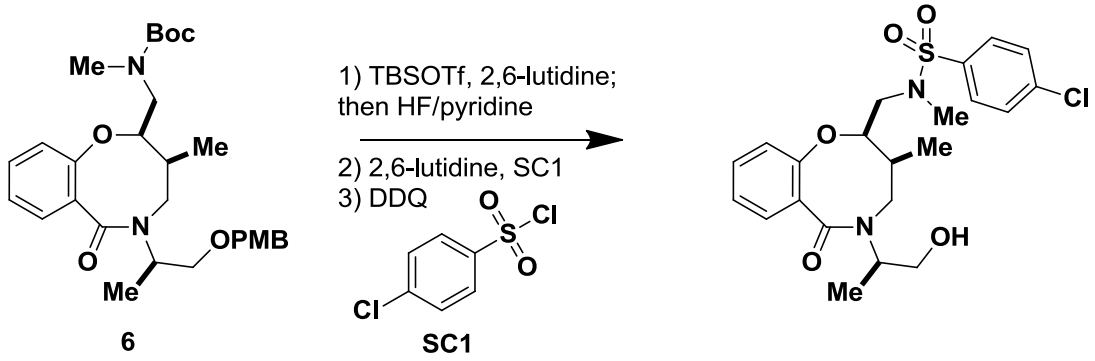
4a **6**

4a (1.0 eq) in EtOH (0.05 M) was stirred with palladium (10% on activated carbon, 0.10 eq) at 35°C under H_2. After the starting material was completely consumed (as monitored by LC-MS, 1-2 h), the mixture was cooled, filtered through celite, and

concentrated. The product was dissolved in a 7:1 mixture of EtOH and acetic acid (final concentration 0.007 M). Aqueous solutions of sodium nitrite (14.0 eq) and sodium bisulfite (10.0 eq) were added. The reaction was stirred at room temperature and monitored by LCMS. After 20 h, an additional 7.0 eq of sodium nitrite and 5.0 eq of acetic acid were added and the reaction was stirred at room temperature overnight. The reaction was quenched with water, extracted with chloroform, and concentrated. The product was purified using flash chromatography on silica gel (10-30% EtOAc in hexanes).

6: Following the general protocol, 1.06 g of **4a** was reacted with 0.20 g palladium (10% on activated carbon) in 38 mL EtOH. 0.13 g of the resulting product was reacted with 0.24 g (14 eq) $NaNO_2$ and 0.26 g (10 eq) $NaHSO_3$ and later again with 0.12 g (7 eq) $NaNO_2$ and 0.13 (5 eq) $NaHSO_3$ in 31 mL EtOH and 4.4 mL AcOH (7:1 mixture) to give **6** 48% yield.

tert-butyl-(((2R,3S)-5-((R)-1-((4-methoxybenzyl)oxy)propan-2-yl)-3-methyl-6-oxo-3,4,5,6-tetrahydro-2H-benzo[b][1,5]oxazocin-2-yl)methyl)(methyl)carbamate **6**: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 7.51 (d, J = 6.5 Hz, 1H), 7.29 – 7.18 (m, 3H), 6.96 – 6.79 (m, 4H), 4.71 – 4.57 (m, 2H), 4.52 – 4.42 (m, 2H), 3.79 (s, 3H), 3.76 – 3.66 (m, 1H), 3.58 – 3.49 (m, 1H), 3.46 – 3.33 (m, 2H), 3.28 – 3.17 (m, 2H), 2.79 (s, 1.5H), 2.73 (s, 1.5H), 1.40 (s, 9H), 1.32 – 1.22 (m, 3H), 0.94 (d, J = 6.6 Hz, 3H), 0.89 – 0.81 (m, 1H). LRMS-ES$^+$ *m/z*: 513.31 (M^+).



tert-butyl-(((2R,3S)-5-((R)-1-((4-methoxybenzyl)oxy)propan-2-yl)-3-methyl-6-oxo-3,4,5,6-tetrahydro-2H-benzo[b][1,5]oxazocin-2-yl)methyl)(methyl)carbamate **6** (1.0 eq) was dissolved in DCM (0.1 M) and 2,6-lutidine (4.0 eq) and TBSOTf (3.0 eq) were added at room temperature. The reaction was stirred for 2h and subsequently quenched with saturated NH_4Cl solution. The mixture was extracted with EtOAc and the combined organic extracts were dried over $MgSO_4$, filtered, and concentrated to produce the crude silyl carbamate which was dissolved again in THF (0.1 M). HF/pyridine (70%, 1.0 eq) was added and the mixture was stirred for 40 min at room temperature. The reaction was then quenched with saturated NH_4Cl solution and extracted with EtOAc. The combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The crude material was dissolved in DCM (0.03 M), and 2,6-lutidine (6.0 eq) was added and the mixture was cooled to 0°C. The desired sulfonyl chloride (3.0 eq) was added in minimal DCM. The reaction was stirred at 0°C overnight and subsequently quenched with water. The mixture was extracted with DCM, dried over $MgSO_4$, filtered, and concentrated. The crude product was then dissolved in a 4:1 DCM and pH 7 buffer solution (final concentration: 0.08 M). The mixture was cooled to 0°C and DDQ (1.5 eq) was added. The reaction was stirred at 0° for another 10 mins and then brought to room temperature and stirred overnight. The reaction was quenched with water and extracted

with DCM. The combined organic extracts were washed with saturated $NaHCO_3$ solution and subsequently, activated carbon was added. The mixture was filtered through celite and the filtrand was washed with hot DCM several times. The filtrate was concentrated and the product was purified by flash chromatography on silica gel (0-100% EtOAc in hexanes).

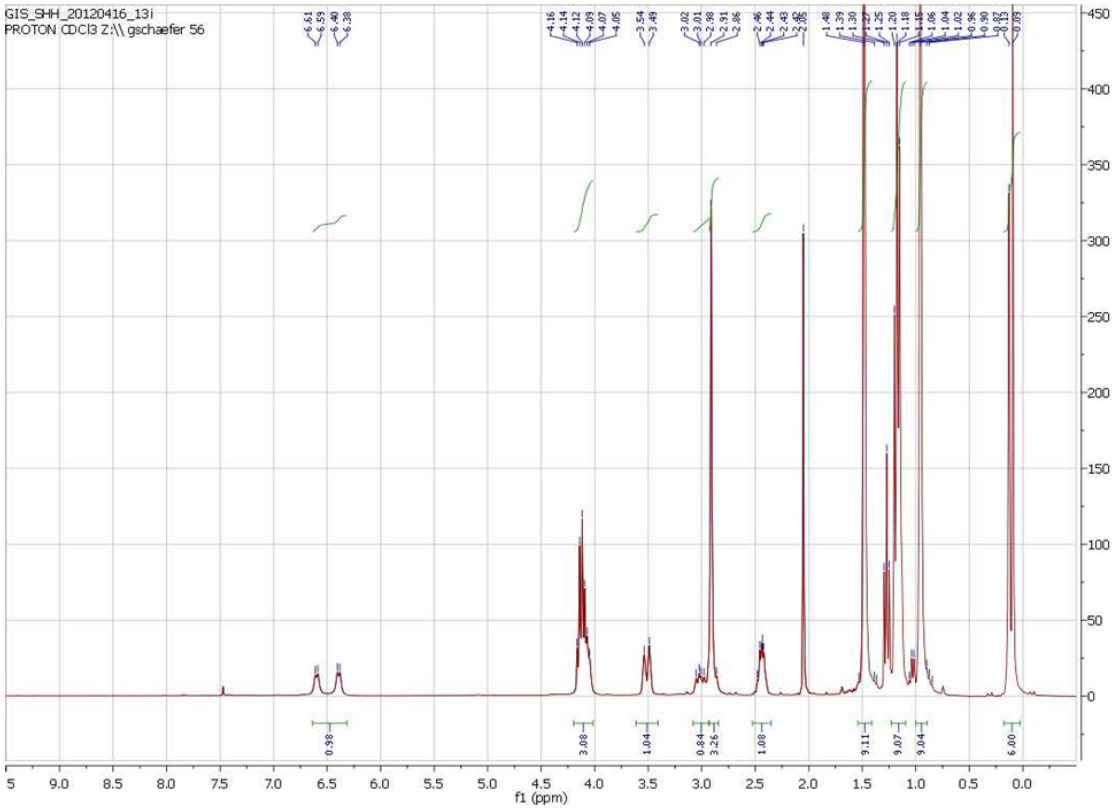
19

Following the general protocol, 0.06 g of **6** was reacted with 0.08 mL TBSOTf and 0.06 mL 2,6-lutidine in 1.2 mL DCM, and later with 11 μL HF/pyridine in 1.2 mL THF. The product was reacted with 0.7 g **SC1** and 0.08 mL 2,6-lutidine in 4.7 mL DCM. The resulting product was reacted with 0.04 g DDQ in 4 mL DCM and 1 mL pH 7 buffer solution to give **19** in 49% overall yield.

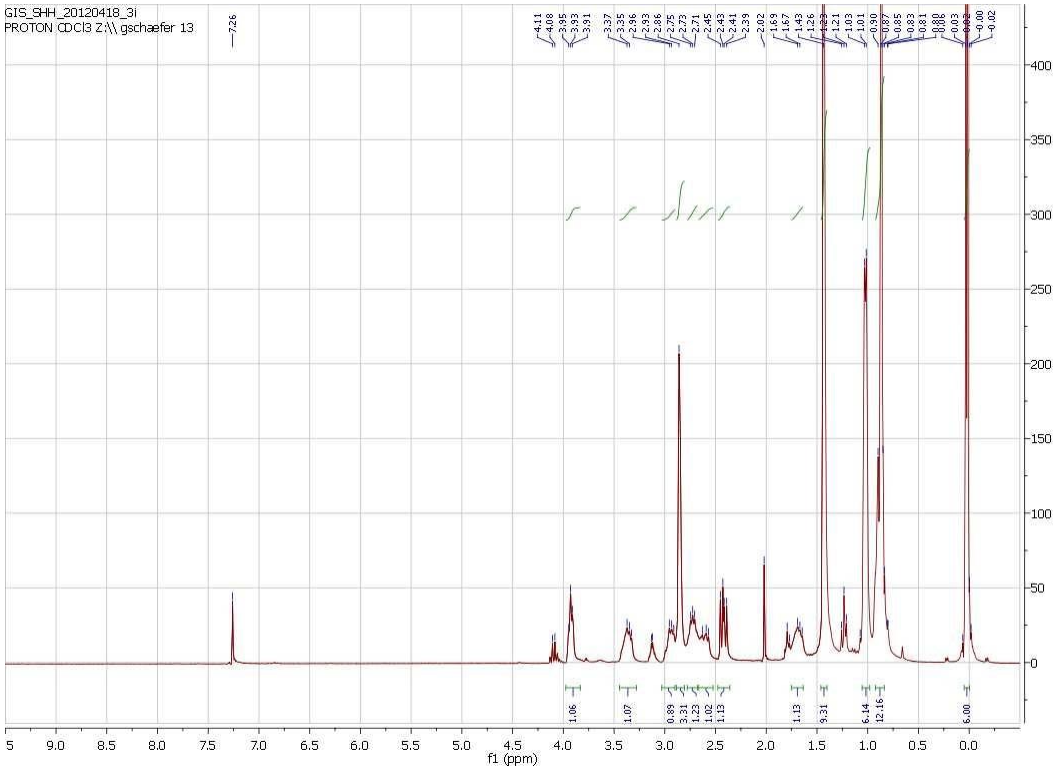
19: ^{1}H NMR (300 MHz, $CDCl_3$, room temperature) δ 7.70 (d, *J* = 8.5 Hz, 2H), 7.56 – 7.34 (m, 3H), 7.32 – 7.16 (m, 1H), 6.95 (t, *J* = 7.5 Hz, 1H), 6.76 (d, *J* = 8.3 Hz, 1H), 4.88 – 4.76 (m, 1H), 4.74 – 4.56 (m, 1H), 3.90 (d, *J* = 11.4 Hz, 1H), 3.73 – 3.58 (m, 1H), 3.53 (dd, *J* = 13.9, 5.2 Hz, 1H), 3.36 – 3.23 (m, 2H), 3.08 (dd, *J* = 13.9, 7.4 Hz, 1H), 2.81 (s, 3H), 2.28 – 2.15 (m, 1H), 1.97 – 1.85 (m, 1H), 1.29 (d, *J* = 7.0 Hz, 3H), 1.01 (d, *J* = 6.6 Hz, 3H). HRMS (ESI) calcd for $C_{22}H_{28}ClN_2O_5S$ $[M+H]^+$: 467.1407. Found: 467.1409.

Spectra

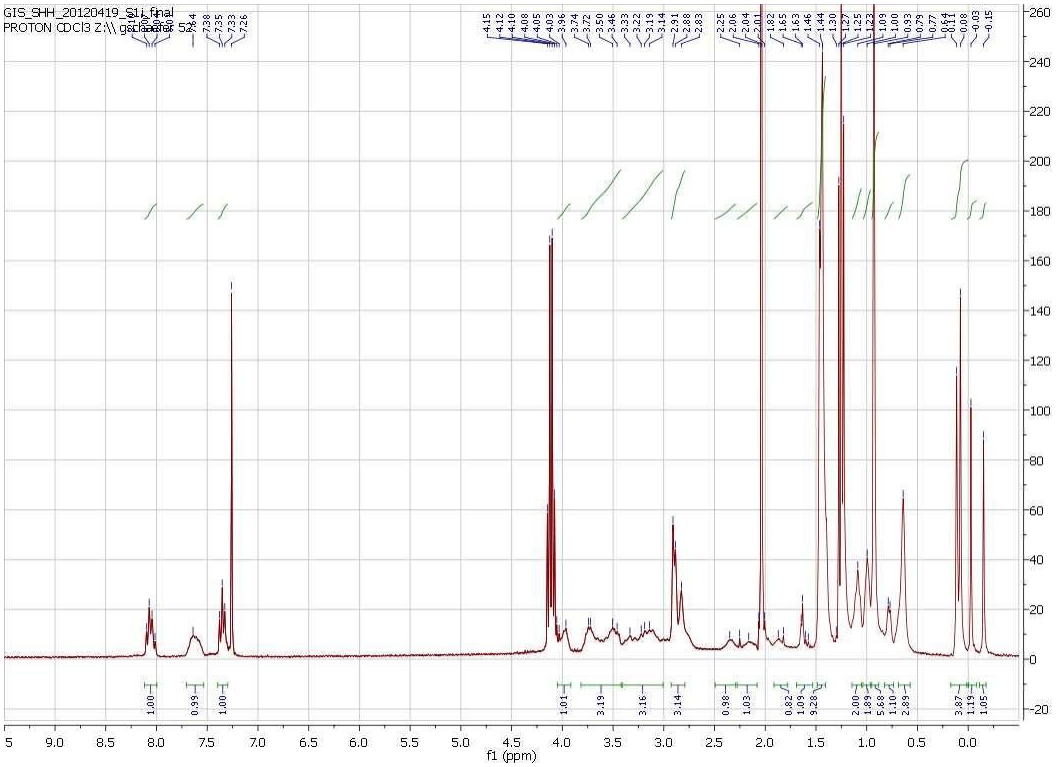
I1b ^{1}H NMR:



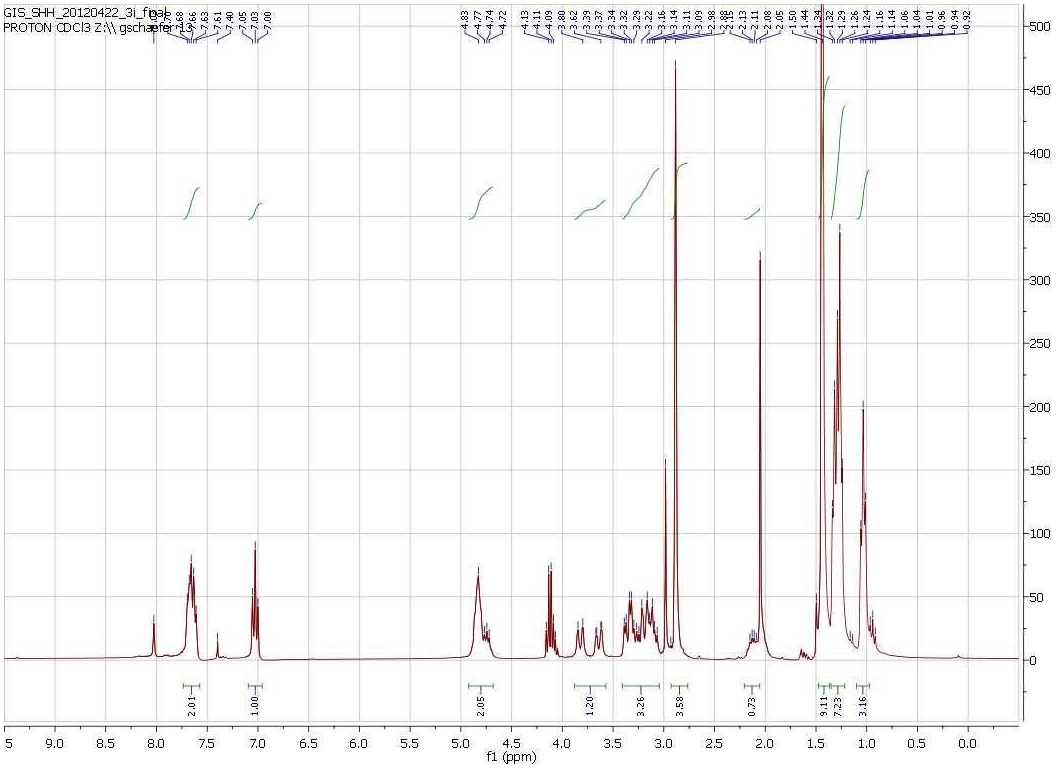
3b ^{1}H NMR:



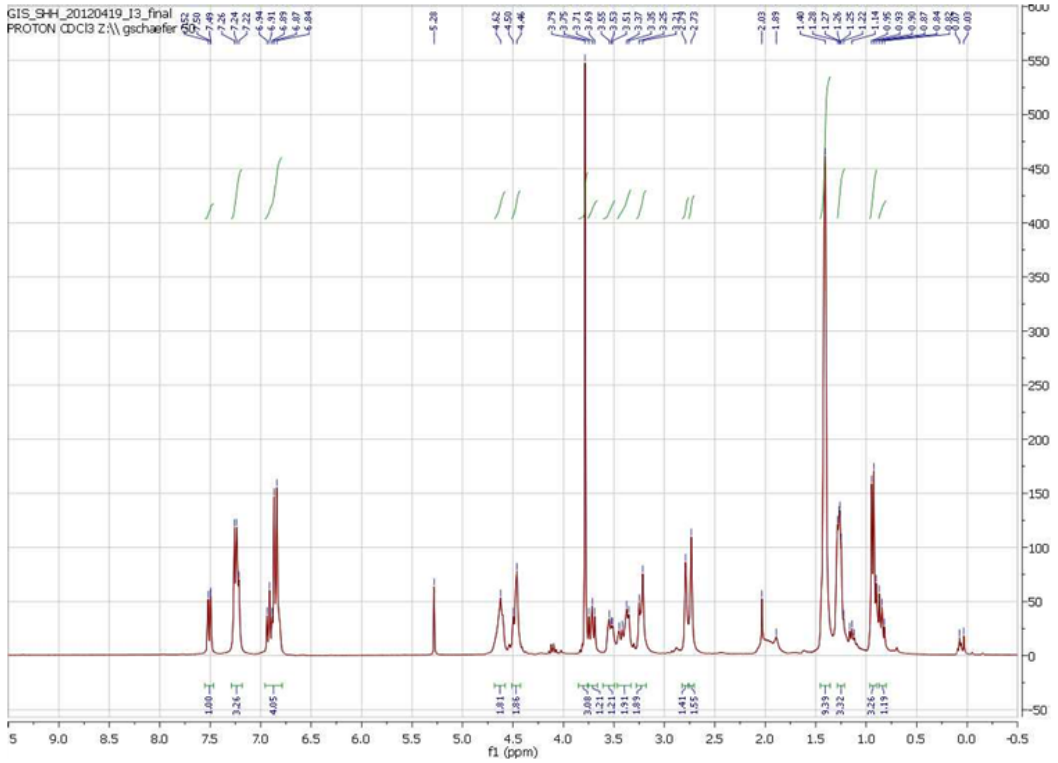
I4b ^{1}H NMR:



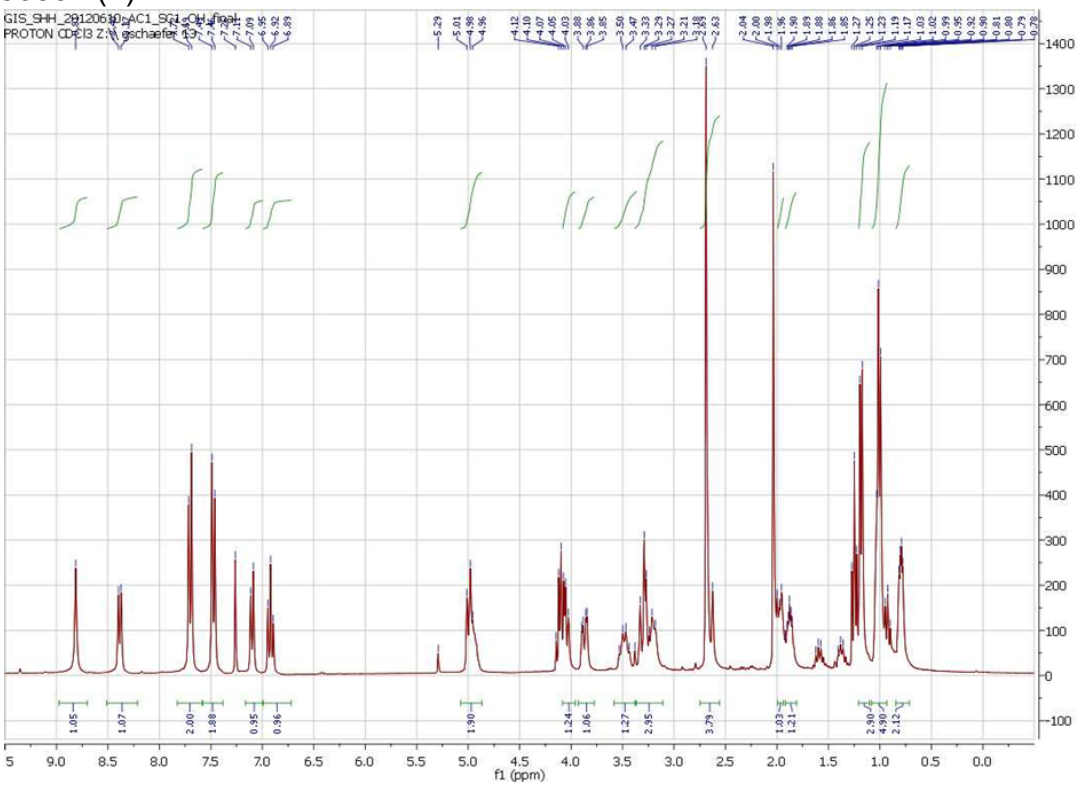
4b ^{1}H NMR:



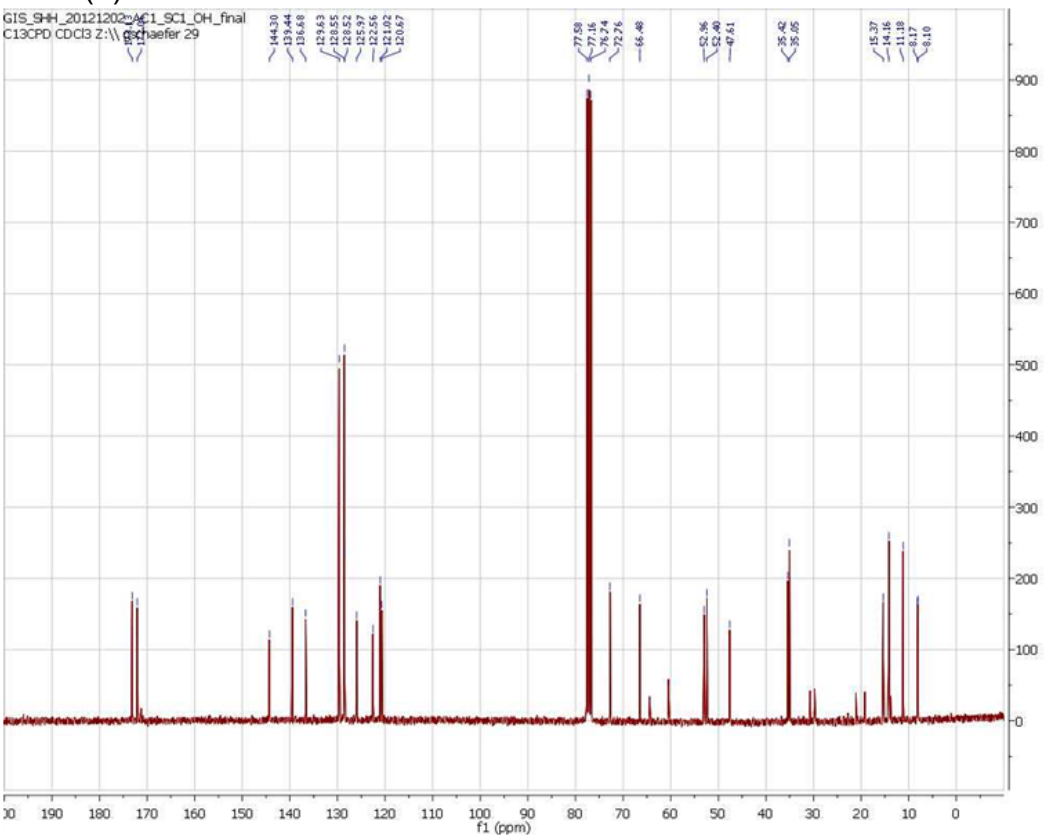
6 ^{1}H NMR:



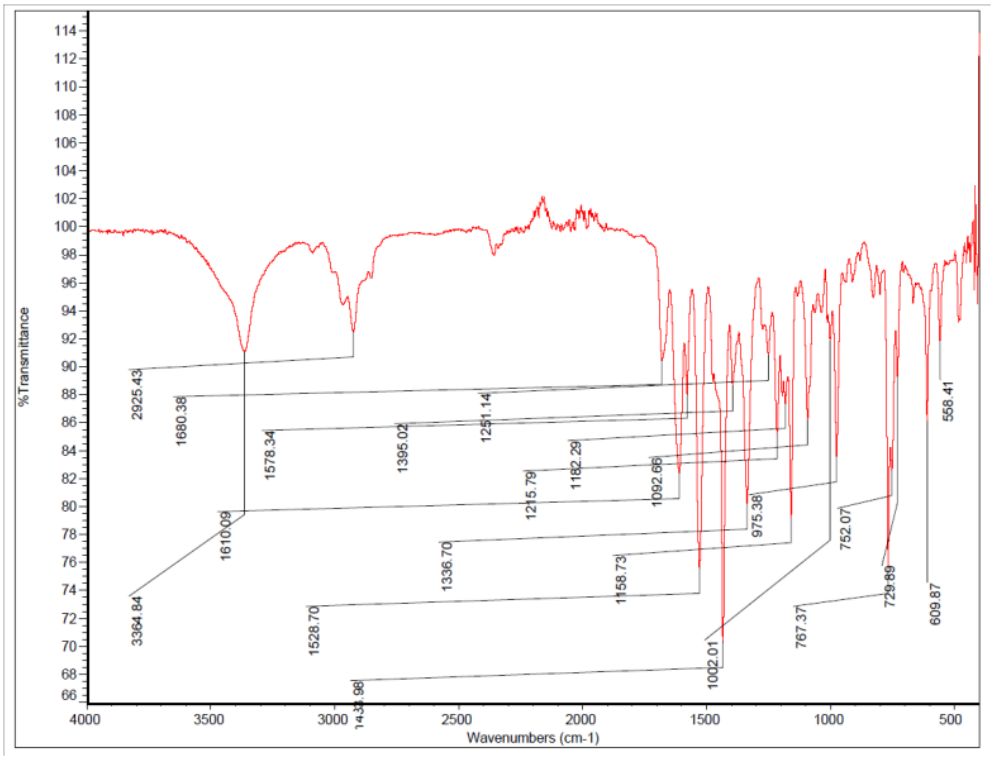
BRD50837 (**7**) ^{1}H NMR:



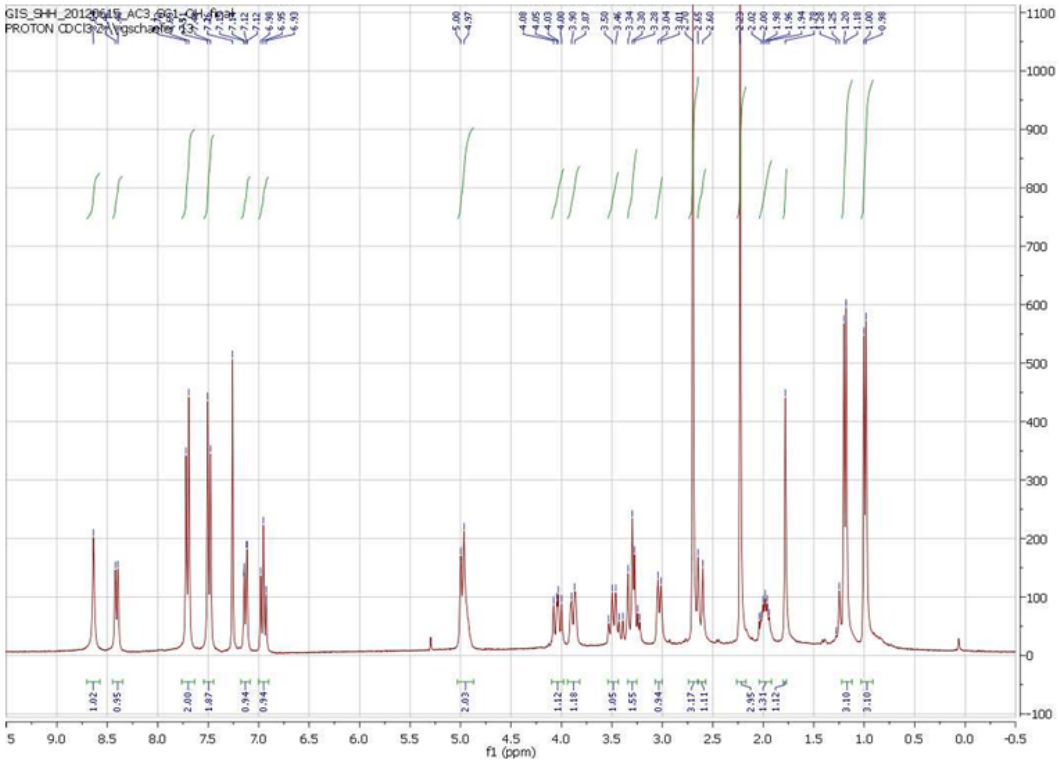
BRD50837 (**7**) ^{13}C NMR:



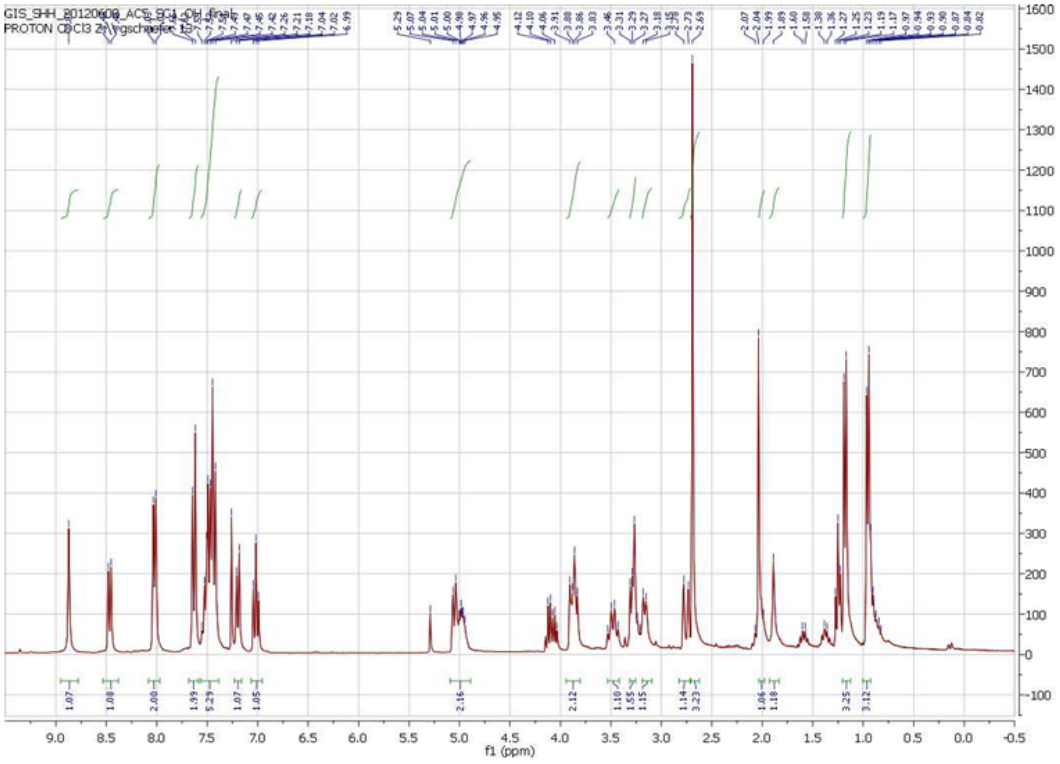
BRD50837 (**7**) IR:



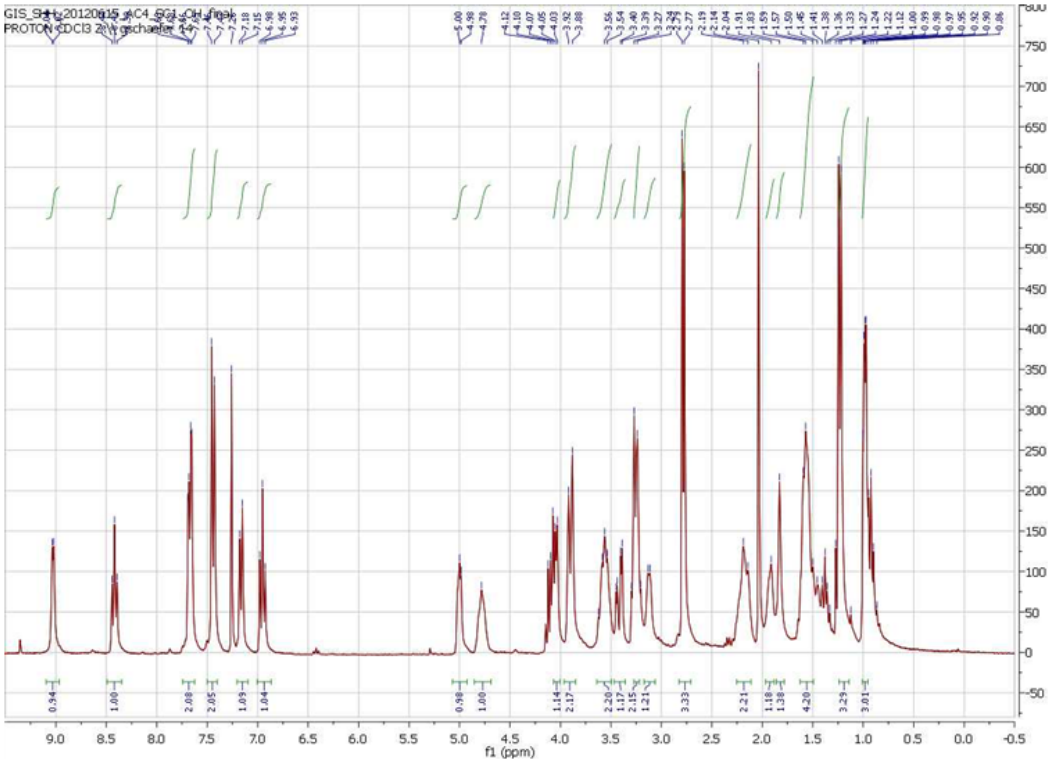
8 ^{1}H NMR:



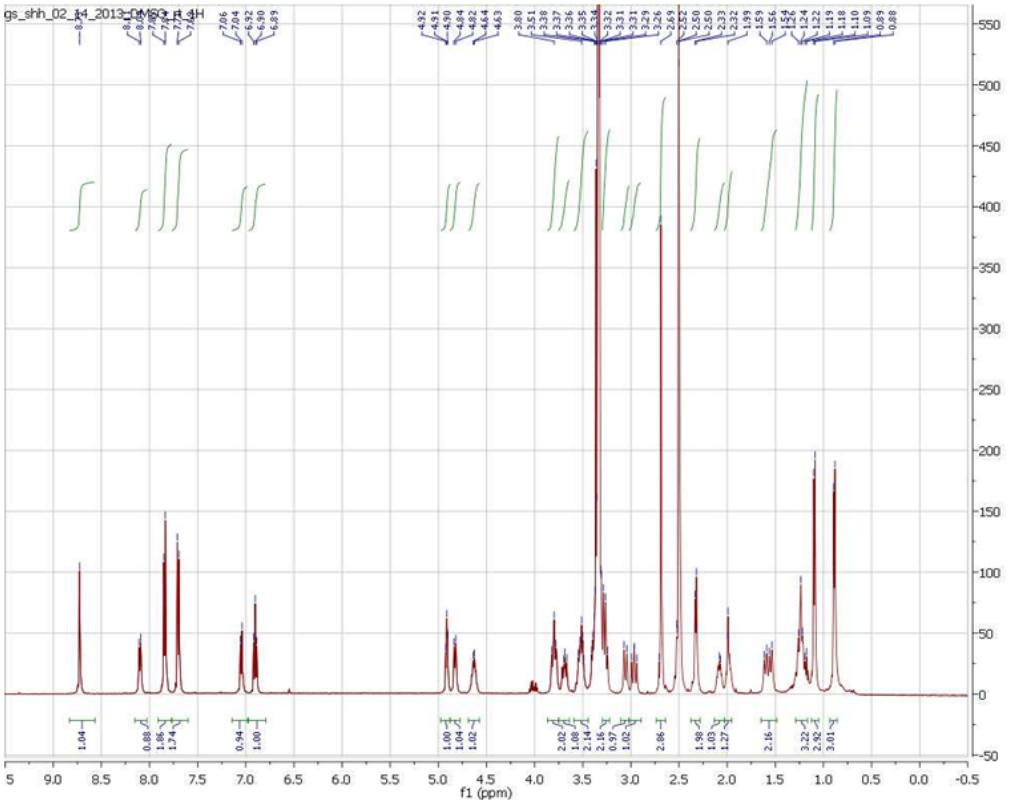
9 ^{1}H NMR:



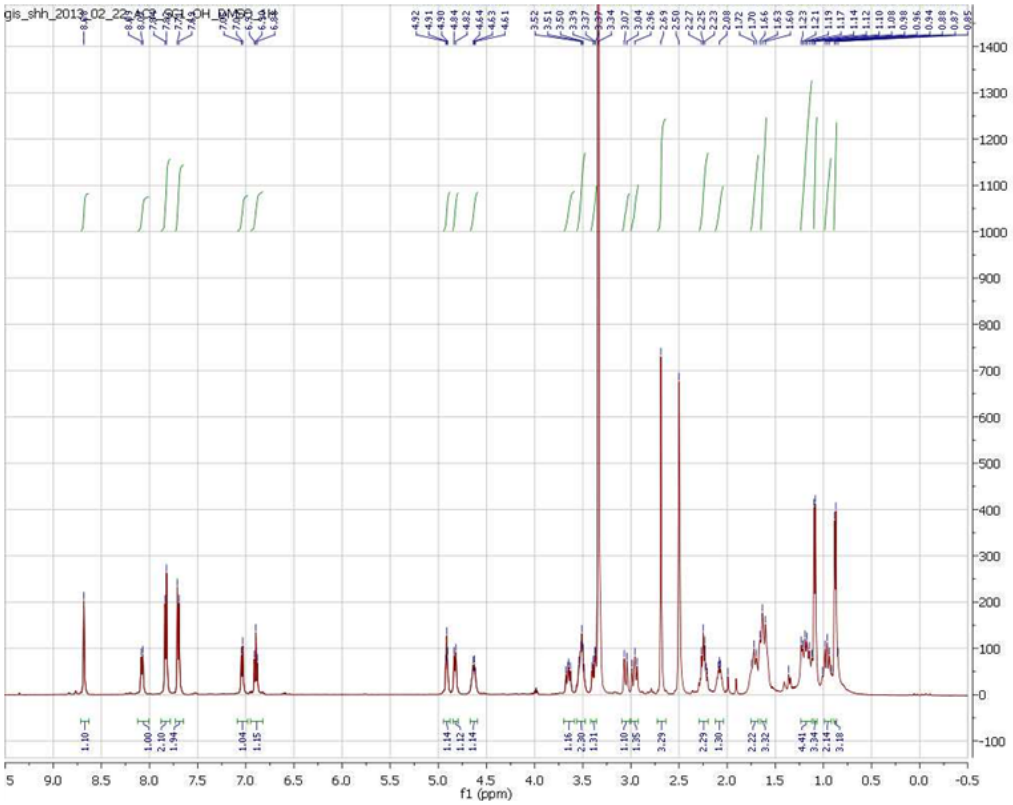
10 ^{1}H NMR:



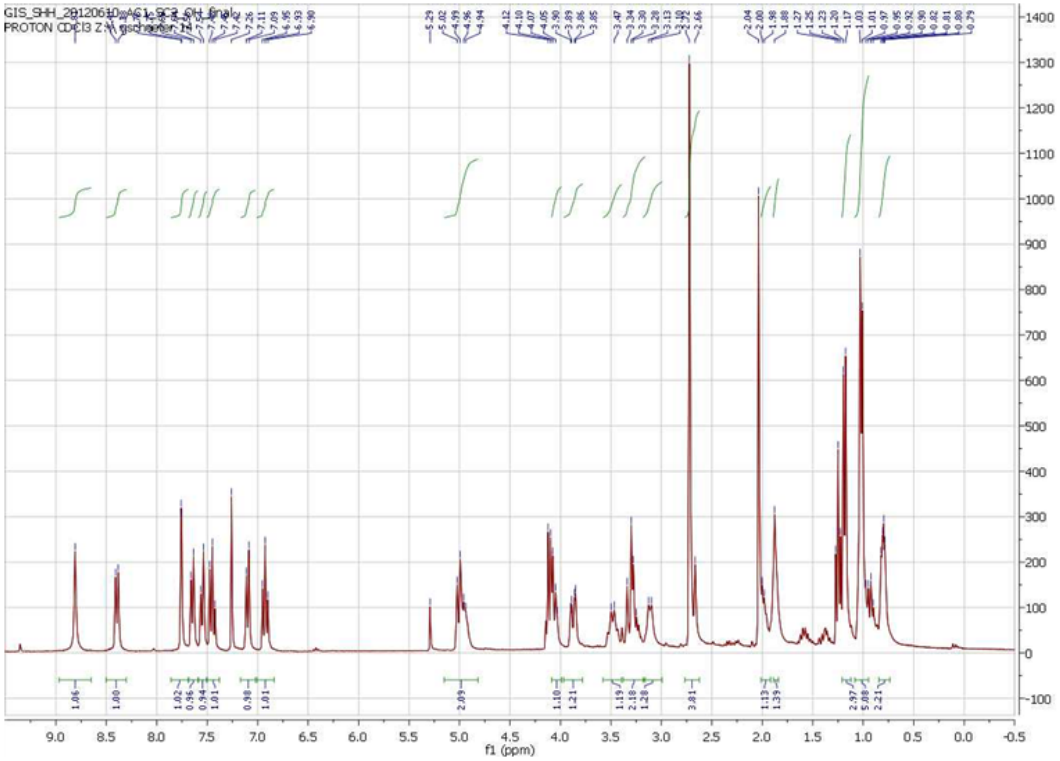
11 ^{1}H NMR:



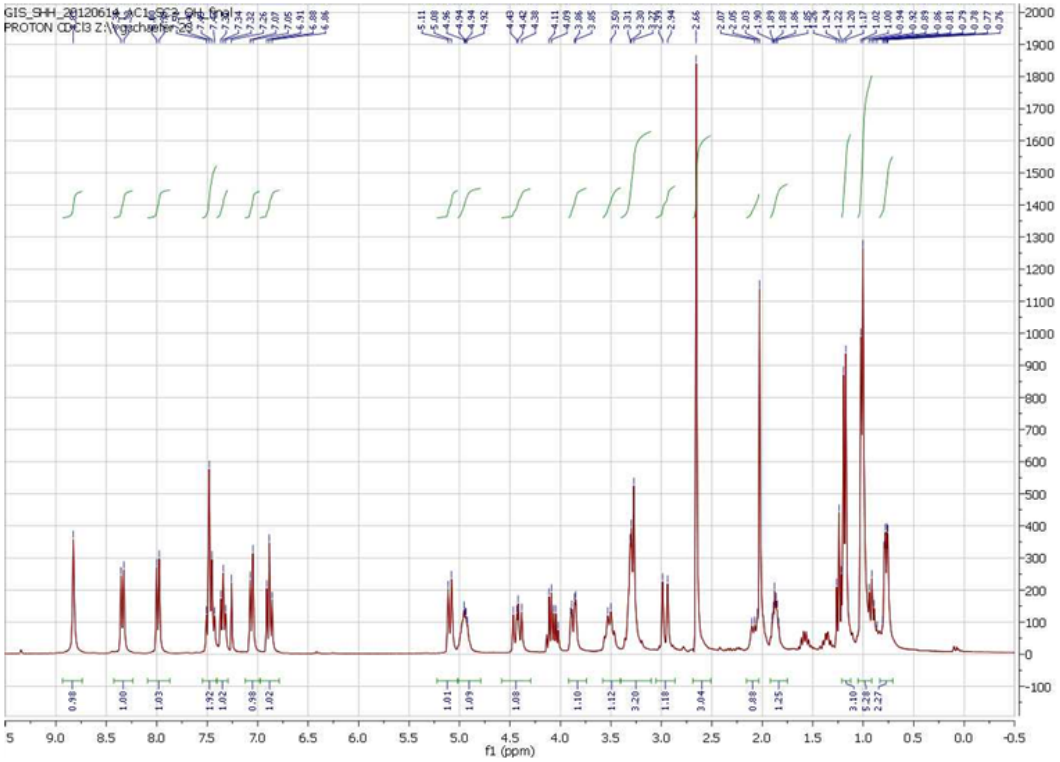
12 ^{1}H NMR:



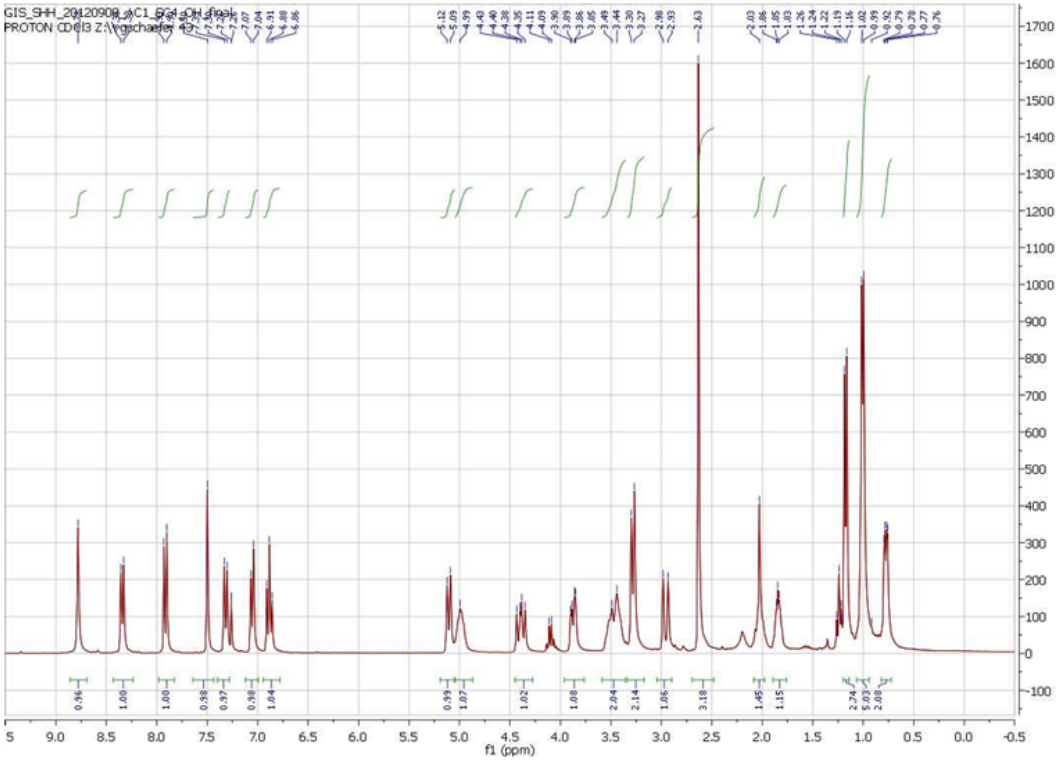
13 ^{1}H NMR:



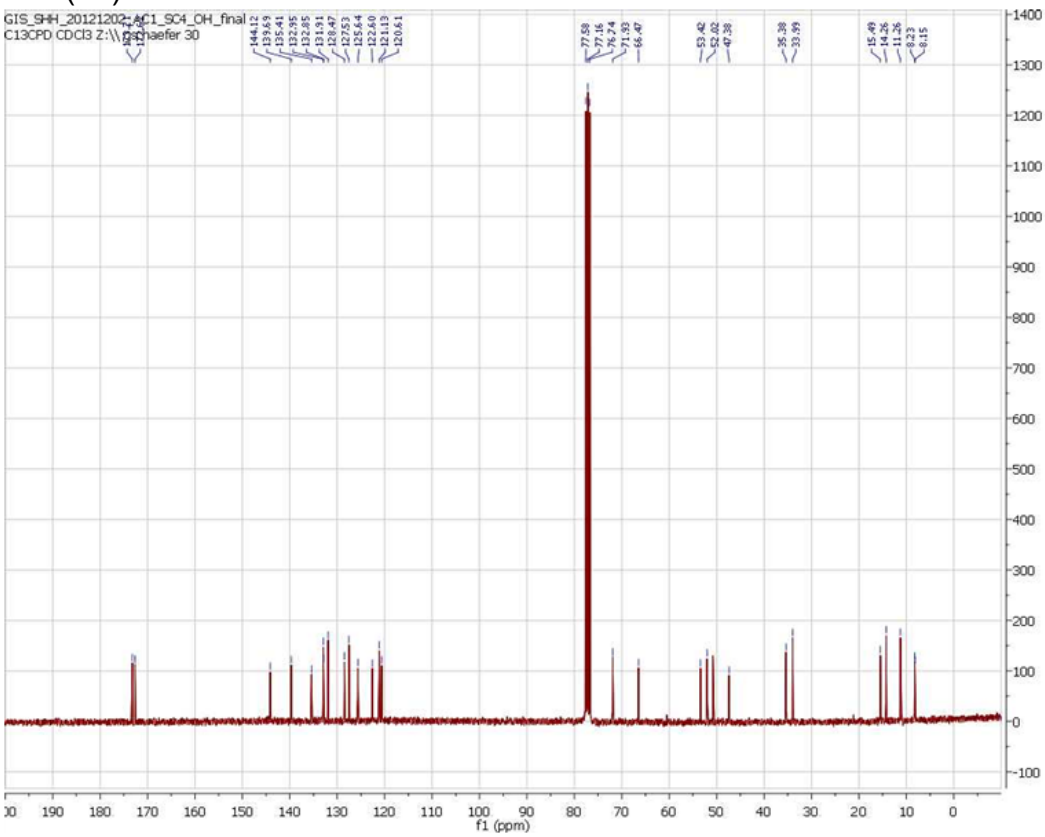
14 ^{1}H NMR:



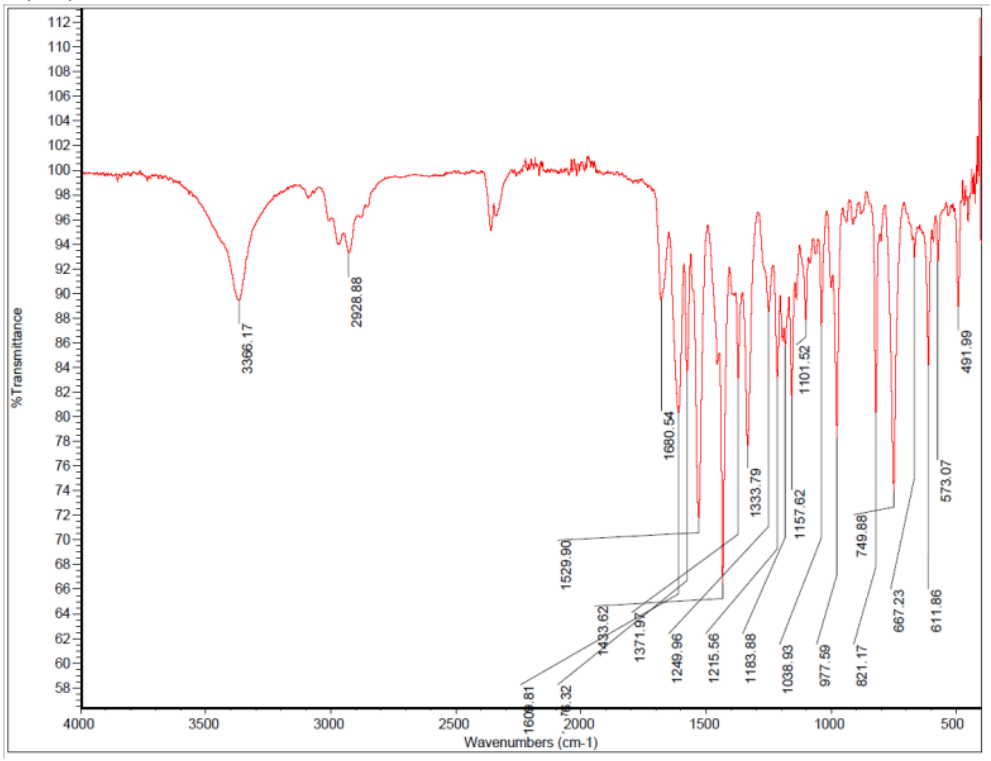
BRD9526 (**15**) ^{1}H NMR:



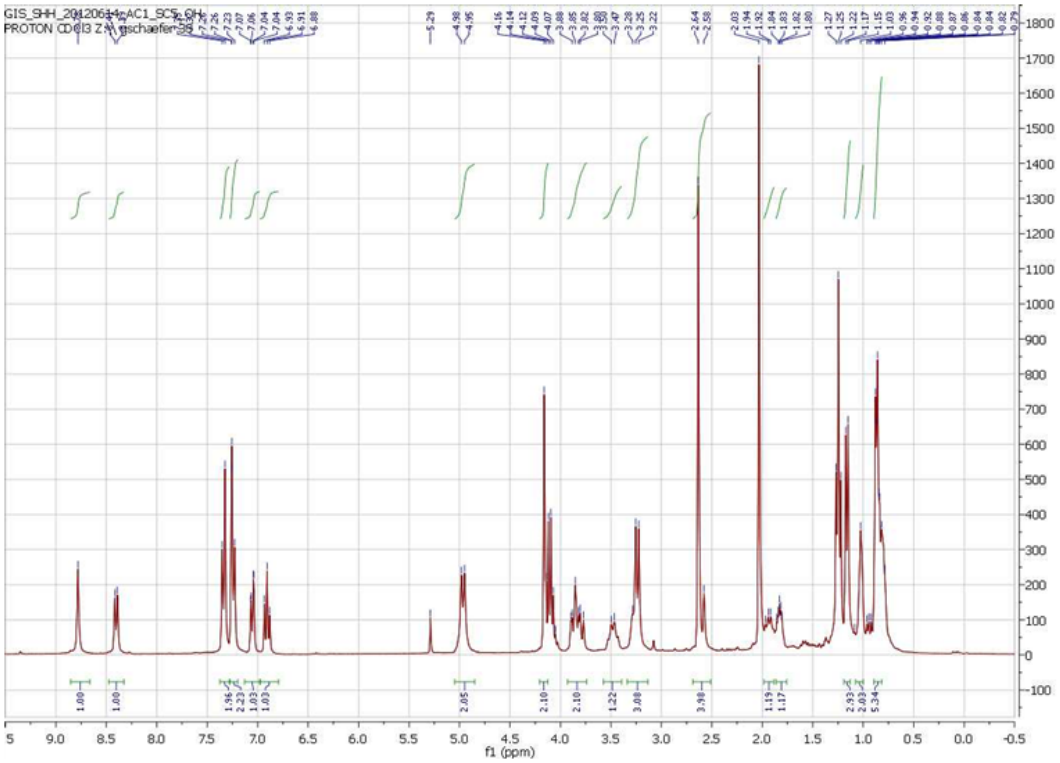
BRD9526 (**15**) ^{13}C NMR:



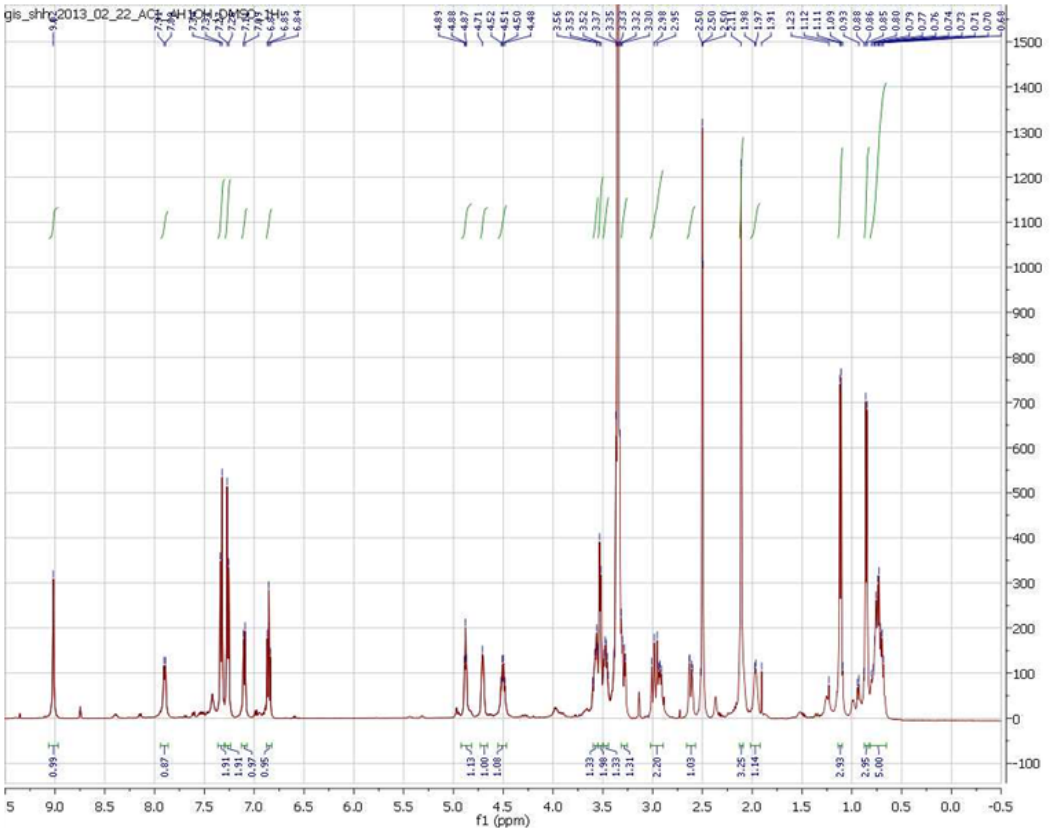
BRD9526 (**15**) IR:



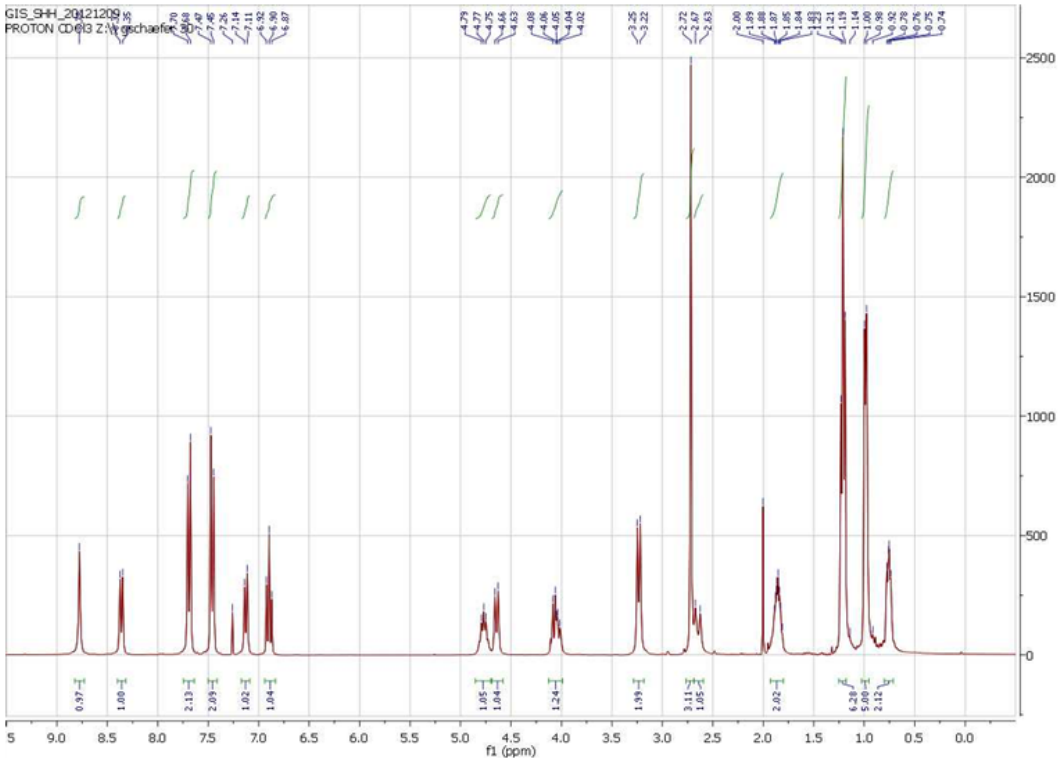
16 ^{1}H NMR:



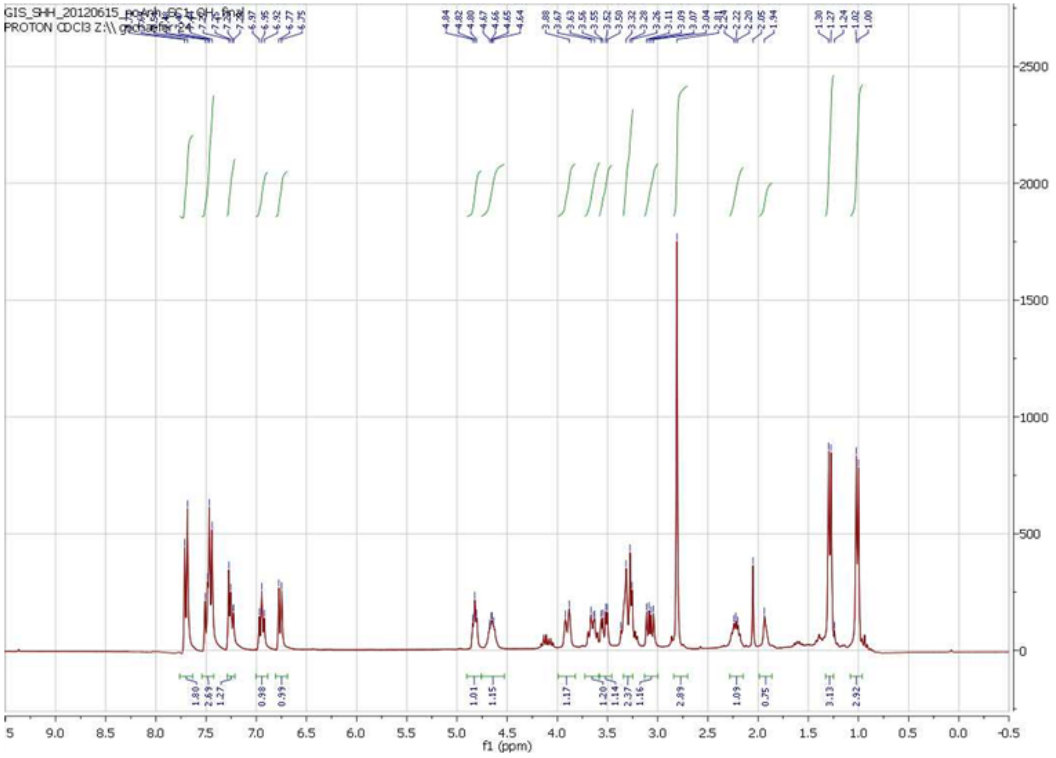
17 ^{1}H NMR:



18 ^{1}H NMR:



19 ^{1}H NMR:



References

(1) Reprinted (adapted) with permission from Schaefer, G. I. *et al.* "Discovery of Small-Molecule Modulators of the Sonic Hedgehog Pathway." *J. Am. Chem. Soc.*, *135*, 9675-9680. Copyright (**2013**) American Chemical Society.

(2) Marcaurelle, L. A. *et al.* "An aldol-based build/couple/pair strategy for the synthesis of medium- and large-sized rings: discovery of macrocyclic histone deacetylase inhibitors." *J. Am. Chem. Soc.* **2010**, *132*, 16962-16976.

(3) Chou, D. H. *et al.* "Synthesis of a novel suppressor of beta-cell apoptosis via diversity-oriented synthesis." *ACS Med. Chem. Lett.* **2011**, *2*, 698-702.

(4) Gottlieb, H. E., Kotlyar, V. & Nudelman, A. "NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities." *J. Org. Chem.* **1997**, *62*, 7512-7515.

(5) Livak, K. J. & Schmittgen, T. D. "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." *Methods* **2001**, *25*, 402-408.

Chapter Four

Mechanism-of-action studies on small-molecule modulators of the Sonic Hedgehog pathway

I assayed the compounds' mechanism of action in SAG-induced C3H10T1/2 cells, $Ptch^{-/-}$ cells, and $Sufu^{-/-}$ cells. I prepared and characterized the immobilized analog and its intermediates. I prepared the samples used for iTRAQ analysis together with Steven Poynter. Dr. Monica Schenone and Emily Hartmann processed the samples and ran the iTRAQ Mass-Spectrometry and analysis.

In part adapted from Schaefer, G.I.; Perez, J.R.; Duvall, J.R.; Stanton, B.Z.; Shamji, A.F.; Schreiber, S.L. "Discovery of Small-Molecule Modulators of the Sonic Hedgehog Pathway" *J. Am. Chem. Soc.* **2013**, *135*, 9675-9680.[1]

BRD50837's and BRD9526's level of activity in the Hedgehog pathway

To understand the mechanisms of action of the compounds discovered in Chapter 3, I tested the response to BRD50837 and BRD9526 in several epistasis experiments and in a competition assay in comparison to cyclopamine, a commonly used Smo inhibitor. I first tested the compounds in C3H10T1/2 cells that were treated with SAG (a small-molecule activator of Smo), rather than Shh-conditioned medium, to activate the Hh pathway (Figure 4.1a).[2] Like cyclopamine, the compounds suppressed SAG-induced differentiation, suggesting a mechanism of action involving modulation of a step in the signaling cascade at or following Smo signaling. In parallel, I tested the compounds in $Ptch^{-/-}$ cells, mouse embryonic fibroblasts that contain a *β-galactosidase* reporter gene instead of the *Ptch* gene downstream of the *Ptch* promoter.[3] Lacking the repressor Ptch, the Hh pathway is constitutively active in these cells. In this assay, cyclopamine maintained its inhibition of the pathway, but BRD50837 and BRD9526 had no effect (Figure 4.1b). When viewing the overall pathway as a linear set of response nodes, these results are apparently in contrast to the previous observation – they suggest that the compounds act at the level of Ptch or a step upstream of Ptch signaling. However, I also identified compounds from the original screen with responses similar to cyclopamine in $Ptch^{-/-}$ cell-based, SAG/C3H10T1/2 cell-based, and BODIPY-cyclopamine displacement assays (Chapter 2), which gave me confidence that the assays accurately measure compound/activity profiles.

With these puzzling results in hand, I performed two addtitional assays to characterize the compounds. I first tested the compounds in a competition assay to determine whether they displace BODIPY-cyclopamine in a cellular assay, thus

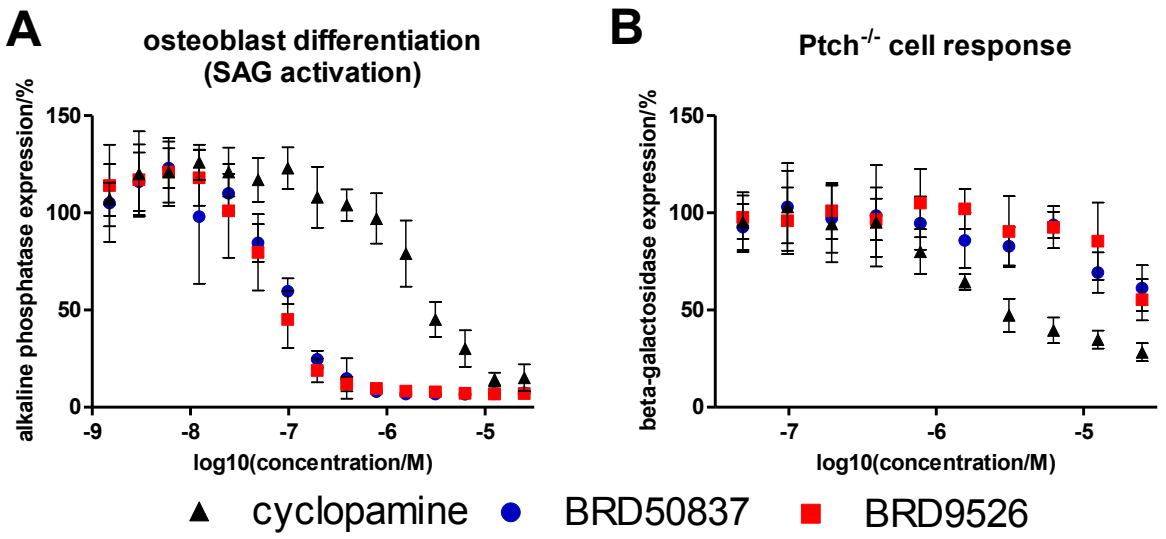


Figure 4.1. All values are shown and generated from three independent experiments run in duplicate (values are calculated average ± SD). a) Inhibition of SAG-induced differentiation of C3H10T1/2 cells by BRD50837, BRD9526, and cyclopamine after 48 h. b) β-galactosidase expression response of $Ptch^{-/-}$ cells to 48 h treatment with BRD50837, BRD9526, and cyclopamine.

suggesting that they bind Smo in the cyclopamine-binding site.[4] Unlike cyclopamine, both BRD50837 and BRD9526 did not lead to a reduction of BODIPY-cyclopamine binding (Figure 4.2a), suggesting that BRD50837 and BRD9526 do not interfere with cyclopamine binding.

I next tested the activity of BRD50837 and BRD9526 in $SuFu^{-/-}$ cells. These mouse embryonic fibroblasts lack the pathway repressor SuFu, which leads to constitutively active Hh signaling.[5] It has been reported that Smo antagonists do not inhibit this signaling while the pathway inhibitor GANT-61 does.[6] In our experiments, cyclopamine partially inhibited downstream *Gli1* expression (Figure 4.2b), perhaps due to an off-target activity observed at high concentrations.[7,8] However, another more potent Smo inhibitor (vismodegib)[9] showed no suppression of *Gli1* expression,

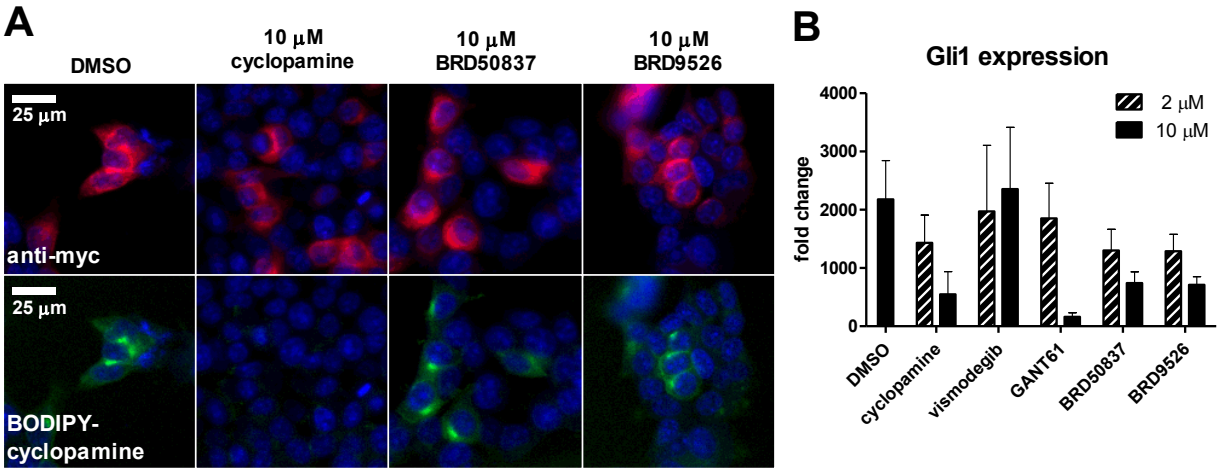


Figure 4.2. a) Effects of cyclopamine, BRD50837, and BRD9526 on BODIPY-cyclopamine (10 nM) binding to exogenously expressed Smo. blue (Hoechst 33342), red (anti-myc), green (BODIPY-cyclopamine). b) Gli1 expression in Sufu$^{-/-}$ cells treated with GANT61, cyclopamine, BRD50837, and BRD9526. All values are shown and generated from three independent experiments run in duplicate (values are calculated average + SD).

consistent with the existing model of SuFu being downstream of Smo (Figure 4.2b). BRD50837 and BRD9526, like cyclopamine, partially lowered *Gli1* expression at concentrations of 2 μM and 10 μM (Figure 4.2b). This partial inhibition may reflect an off-target effect at high concentrations, but it is also possible that these compounds act in a way that influences the pathway at the level of or downstream of SuFu signaling.

The compounds therefore act similarly to cyclopamine, a well-characterized pathway inhibitor, in some aspects (SAG/C3H10T1/2 cell-based and SuFu$^{-/-}$ cell-based assays), but seem to have a different mechanism of action in other aspects (Ptch$^{-/-}$ cell-based and BODIPY-cyclopamine displacement assays). These data suggest that BRD50837 and BRD9526 may function by mechanisms of action that are distinct from cyclopamine and not easily described by traditional linear models of the pathway. Consistent with this notion, BRD50837/BRD9526 repressed *Gli1* expression of

C3H10T1/2 cells to a lesser extent than cyclopamine when the compounds were tested at concentrations that yield similar responses in Shh-conditioned medium-induced differentiation of C3H10T1/2 cells (1 μM, and 10 μM, respectively, Suppl. Figure S3.2).

Identifying possible targets - iTRAQ

To find more definite answers about the compounds' mechanism of action, identification of protein binding partner(s) would be optimal. Different methods for target identification such as biochemical, genomic interactions, or computational methods have been developed.[10,11] I decided to employ a quantitative proteomics approach using the isobaric tags for relative and absolute quantification (iTRAQ) technology to gain more insight into a possible target of BRD50837 and BRD9526.

To identify targets using quantitative proteomics, the small molecule is immobilized to a solid support. In a biochemical affinity purification, the immobilized compound is then incubated with lysate and either soluble active molecule, or, as negative control, inactive molecule or vehicle. Subsequently, the solid support is washed, proteins purified in the pull-down are digested, and each sample is tagged using a different iTRAQ tag (Figure 4.3). These tags are designed to be isobaric, but fragment into different reporter ions when using mass spectrometry (MS). The relative abundance of each tag for a peptide can be used to quantify the ratio of the corresponding protein between the different samples.[12] Non-specific binders of the solid support or linker employed will have a ratio between vehicle and soluble active molecule samples that is approximately 1. Specific binders of the immobilized molecule however, will also bind the soluble active molecule, which competes with the immobilized

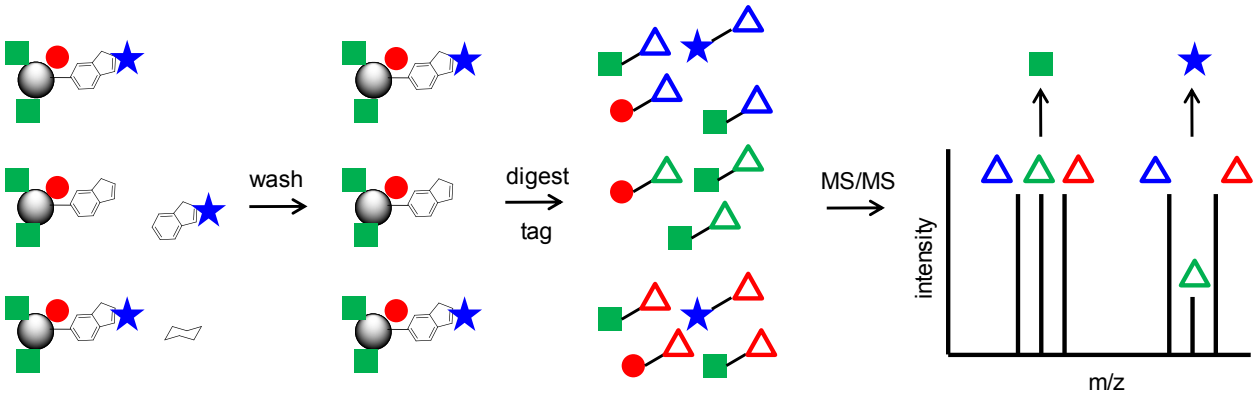
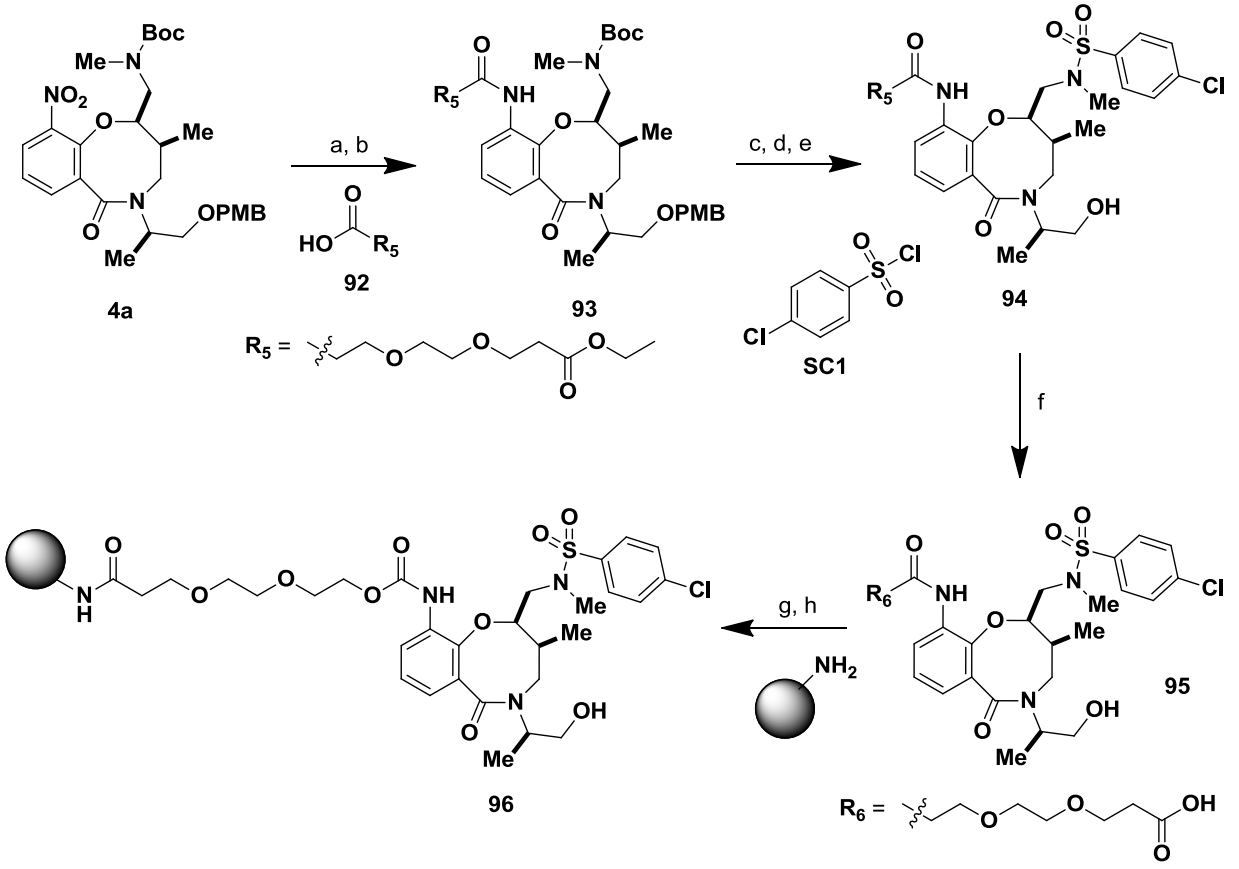


Figure 4.3. Workflow for an iTRAQ experiment. The active small molecule is immobilized on solid support and incubated with cell lysates (solid square, circle, star = proteins). Additionally, the samples will be incubated with either DMSO, soluble competitor (BRD50837), or soluble inactive compound (enantiomer of BRD50837). After washing off any proteins that did not bind the immobilized molecule, the proteins are digested and tagged with iTRAQ tags (triangles). MS/MS quantifies the abundance of each tag for each peptide. Non-specific binders (square) will have similar relative abundance while specific binders (star) will be downregulated in the soluble competitor sample.

molecule. They will get washed off with the soluble competitor, and thus be enriched in the vehicle sample (Figure 4.3).

To prepare the iTRAQ experiment, I first synthesized a derivative of BRD50837 that contained a linker to attach it to the solid support. Because the SAR showed that variations of attachments at the aniline are more easily tolerated, I decided to incorporate a linker at that position. Starting after hydrogenation of scaffold **4a**, the polyethylene glycol linker **92** was added to give **93** (Scheme 4.1). Subsequent deprotection of the Boc group using HF/pyridine, addition of sulfonyl chloride **SC1**, and removal of the PMB with DDQ resulted in compound **94**. This compound was tested for activity in Shh-conditioned medium-induced differentiation of C3H10T1/2. Incorporation of the linker did lead to a loss in activity, but the compound was still very active, but not due to toxicity (Figure 4.4a, 4.4b). Saponification of the ethyl ester (**95**) and activation of

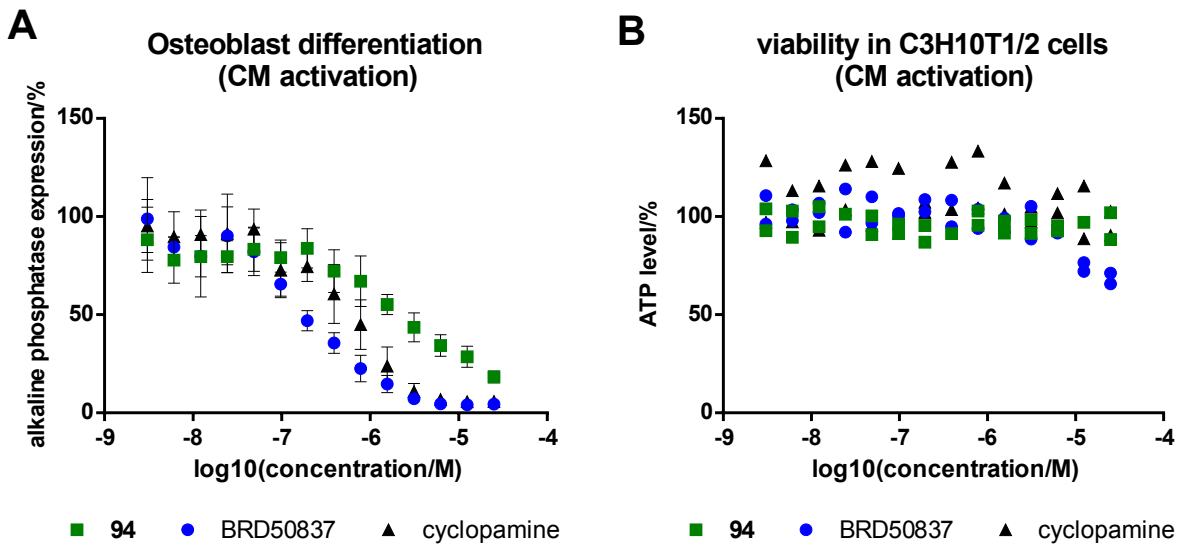
Scheme 4.1. Synthesis of immobilized analog[a]



[a]Reagents and conditions: (a) 10% Pd/C, H_2, EtOH, 35°C. (b) **92**, PyBOP, DIPEA, DCM, rt, 78% over 2 steps. (c) TBSOTf, 2,6-lutidine, DCM, rt, then HF/pyridine, THF, rt. (d) **SC1**, 2,6-lutidine, DCM, 0°C (e) DDQ, pH 7 buffer/DCM, rt, 32% over 3 steps (f) LiOH, 1:1 THF/H_2O, rt, 73%. (g) EDC, NHS, DMF, 40°C. (h) Affigel 102, TEA, DMSO, rt.

the carboxylic acid using EDC and NHS then allowed coupling to the solid support (Scheme 4.1).

For my experiment, I prepared three samples: one using DMSO as a vehicle, one using BRD50837 as a soluble competitor, and one using BRD50837's inactive enantiomer (IE) as a soluble competitor. Specific targets of BRD50837 should be enriched in both the DMSO and the IE samples, but not in the BRD50837 sample (Figure 4.3). The data was analyzed in three groups: group 1 being DMSO/BRD50837, group 2 being IE/DMSO, and group 3 being IE/BRD50837. The two independent



C

Replicate	Proteins
1	1268
2	1298

2+ peptides; 3+ ratios per replicate

Figure 4.4. a) Inhibition of Shh-conditioned medium (CM)-induced differentiation of C3H10T1/2 cells by BRD50837, **94**, and cyclopamine after 48 h. All values are shown and generated from two independent experiments run in duplicate (values are calculated average ± SD). b) Viability of SAG-induced C3H10T1/2 cells in response to 48 h treatment with BRD50837, **94**, and cyclopamine. All values are shown and generated from one experiment run in duplicate. c) number of proteins identified and quantitated in each replicate of the iTRAQ experiment.

experiments run on different days with different iTRAQ tags showed good correlation for overall amount of proteins and peptides purified (Figure 4.4c). However, the histograms of the two replicates show very different distributions of ratios (Figure 4.5 a-c), meaning that the ratios did not replicate well between the two runs and further replicates might be required.

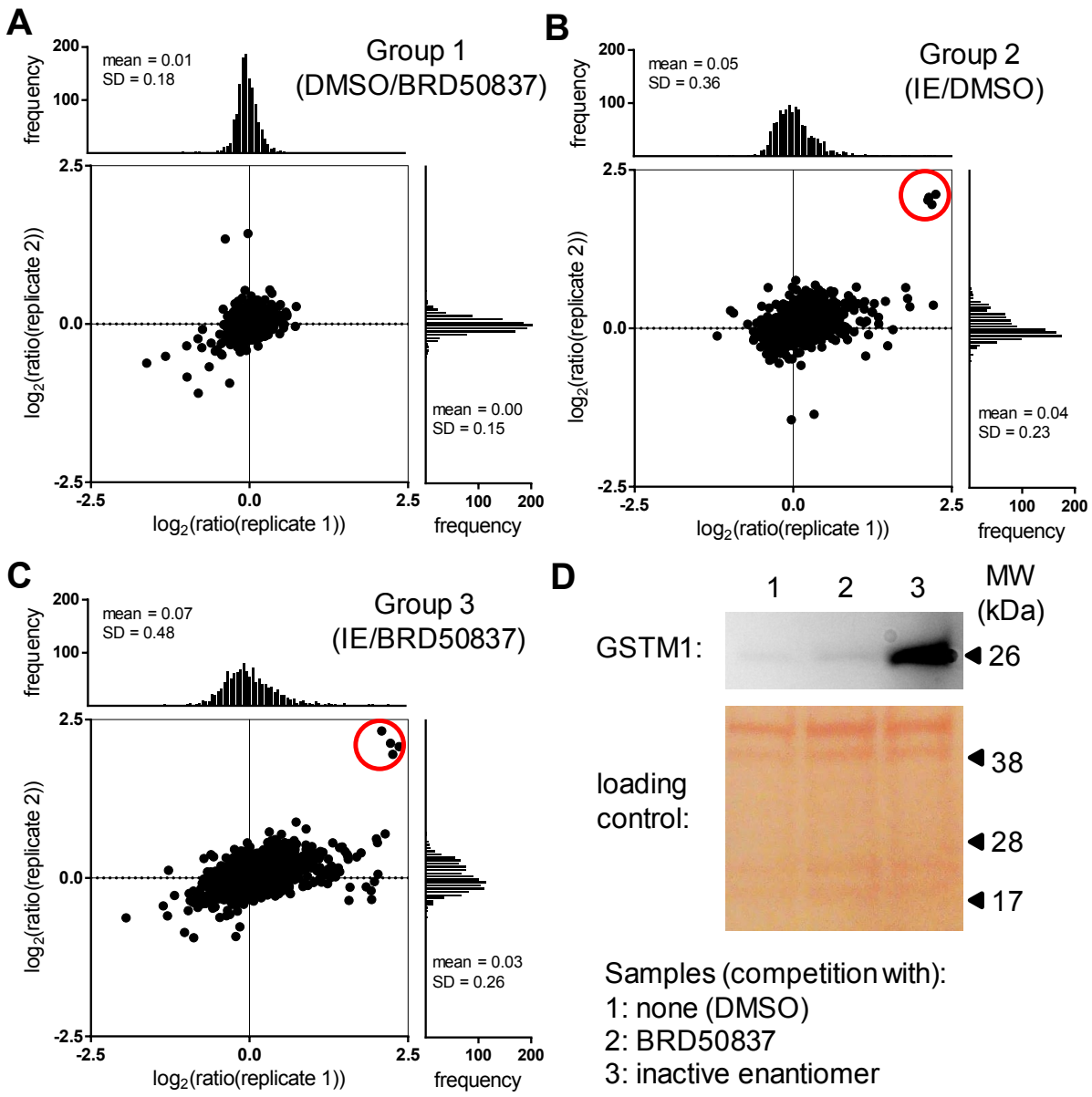


Figure 4.5. a-c) $\log_2$ ratios and frequency histograms for replicate 1 vs. replicate 2 of DMSO/BRD50837 (a), inactive enantiomer (IE)/DMSO (b), IE/BRD50837 (c). Proteins circled in red are isoforms of GSTM. d) western blot analysis for GSTM1 of protein affinity purification. Competition used in lanes: 1 = none (DMSO only), 2 = BRD50837, 3 = BRD50837s inactive enantiomer. Ponceau S stain shows overall protein content in the lanes.

Despite these insufficiencies, there is a group of four proteins that demand attention. These proteins are outliers with a higher signal ratio in Group 2 and Group 3. All four are different proteins from the same family, glutathione S-transferase μ (GSTM1,

GSTM2, GSTM4, GSTM5) (Figure 4.5b, 4.5c). GSTMs are not a protein subfamily frequently detected as targets in other iTRAQ experiments run in different systems,[13] making these outliers of interest. Glutathione S-transferases (GSTs) are involved in drug metabolism. In cells, foreign compounds, such as drugs or probes, are tagged by GSTs through transfer of an abundant cofactor, glutathione. This transfer marks the agents for export from the cells and subsequent excretion.[14,15]

Comparing the IE and BRD50837 samples is the most direct comparison for the competition experiment, as in both cases a small-molecule competitor is in solution, with the difference being the competitor's stereochemistry. Curiously, the GSTMs are not enriched in the DMSO vs. BRD50837 sample (Figure 4.5a). While this could be due to the different environment created by lack of any small-molecule, it might also mean that the enrichment of GSTMs is due to a bias within the experiment. However, a preliminary repetition of the pull-down with subsequent gel electrophoresis and western blot detection of GSTM1, confirms the results from the iTRAQ experiment (Figure 4.5d). In addition, it is striking that four different proteins from the same subfamily, and only that subfamily, showed enriched ratios. GSTs share about 60% sequence identity within a class, but less than 30% between classes, suggesting specificity in binding interactions in the affinity purification.[16] Therefore, these experiments suggest an interaction of BRD50837 and GSTMs in cell lysates, but more experiments are needed to determine the nature of the interaction and possible impacts on the Hh pathway. These could include orthogonal measures of direct binding such as isothermal calorimetry or surface plasma resonance, *in vitro* assessment of enzyme inhibition/activation, and proof of phenotype using RNAi or overexpression of the candidate target.

One other caveat that should be considered with biochemical affinity purifications is the difficulty in identifying membrane proteins or proteins with low cellular abundance as targets.[10,17,18] Additionally, immobilization of the small molecule to the beads can potentially destroy the interactions with the target, even if the compound with linker was still active. Therefore, while GSTMs potentially interact with the compounds in cells and this interaction can be further investigated, additional experiments directed at identifying a target might be necessary. These could include using affinity purification, but performing the pull-down in live cells with covalent capture[18] or immobilization of compounds at different points of attachment. Gene expression profiles could furthermore enhance the understanding of pathways modulated by the compounds.

Conclusion

I report here the discovery of BRD50837 and BRD9526, two selective small-molecule inhibitors of the Shh pathway. Though similar in some respects to traditional pathway inhibitors, the compounds show a distinct pattern of activity in cells perturbed for components of the pathway. The basis for these differences is not yet known, but it hints at the complexity of the pathway. Studies to identify cellular targets have generated GSTMs as potential interacting proteins. Further experiments will be necessary to confirm these interactions with BRD50837 and BRD9526 and effects on the pathway. Elucidating the compounds' mechanism of action will help to realize their full potential as probes and enable the study of this enigmatic pathway.

Experimental Section

Materials and Methods

Organic Synthesis & Compound Characterization: as described in Chapter 3. In addition: Affigel 102 gel as solid support was purchased from Biorad (Hercules, CA, USA). NMR spectra were measured on a Bruker Ultrashield 400 (400 MHz ^{1}H, 100 MHz ^{13}C) (Billerica, MA, USA).

Assay Treatments and Readouts: as described in Chapter 2 & 3. In addition: $SuFu^{-/-}$ and $SuFu^{+/+}$ cells were kindly provided by Dr. Rune Toftgård and Dr. Stephan Teglund (both Karolinska Institutet, Sweden). $SuFu^{-/-}$ and $SuFu^{+/+}$ cells were grown at 37°C/5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) (Cellgro, Manassan, VA, USA) supplemented with 10 % fetal bovine serum (FBS) (Invitrogen, Carlsbad, CA, USA), 2 mM L-glutamine, 1 mM sodium pyruvate (both Life Technologies, Carlsbad, CA, USA), and 10 μg/ml gentamicin. Cells were lysed for iTRAQ using a ModRIPA (50 mM Tris-HCl, pH 7.8, 150 mM NaCl, 0.1% sodium deoxycholate, 1 mM EDTA) buffer with 1% NP-40. Protein concentrations were measured using a Pierce BCA Protein Assay Kit from Thermo Scientific (Waltham, MA, USA) on a Molecular Devices SpectraMax M5 microplate reader (Sunnyvale, CA, USA). iTRAQ reagents were purchased from AB Sciex (Framingham, MA, USA). Samples were read out using a Thermo Scientific Q Exactive Mass Spectrometer. Gels, protein ladder, buffers (loading and running) for gel electrophoresis, and western blotting membranes were purchased from Life Technologies (Carlsbad, CA, USA). Ponceau S solution to visualize protein content in membranes was purchased from Sigma-Aldrich (St Louis, MO, USA). Proteins were transferred from gels to membranes using an Invitrogen iBlot transfer device (Carlsbad,

CA, USA). Milk and 2° antibody for western blotting (anti-rabbit IgG) were purchased from Cell Signaling Technology (Beverly, MA, USA). 1° antibody for western blotting (Rabbit IgG, anti-GSTM1) was purchased from EMD Millipore (Darmstadt, Germany). SuperSignal West Femto Substrate kit was purchased from Thermo Scientific (Waltham, MA, USA). Developed membranes were visualized using a Kodak Carestream Image Station 4000MM Pro (Bruker, Billerica, MA, USA) with chemiluminescence settings for 2 minutes.

Assay Protocols

C3H10T1/2 differentiation assay, Ptch$^{-/-}$ β-galactosidase assay, C3H10T1/2 & Ptch$^{-/-}$ CellTiter-Glo Viability assay, BODIPY-cyclopamine competition assay, RNA extraction, cDNA Reverse Transcription, qPCR: as described in Chapter 2 & 3.

Sufu$^{-/-}$ cell treatment for gene expression: To test the compounds' mechanism of action, they were tested in SuFu$^{-/-}$ mouse embryonic fibroblasts. The SuFu$^{-/-}$ cells and the SuFu$^{+/+}$ cells were seeded with 130,000 cells/well into 6 well plates in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 1 mM sodium pyruvate, and 10 μg/mL gentamicin. The cells were incubated at 37°C/5% CO_2 for 24 hours. Afterwards the wells were washed and DMEM supplemented with 0.5% FBS and 1% penicillin/streptomycin was added. The compounds/DMSO were added in the respective concentrations to the Sufu$^{-/-}$ cells and the respective DMSO amount was added to the SuFu$^{+/+}$ cells. The cells were then incubated for 72 hours at 37°C/5% CO_2. The cells were subsequently harvested by washing them once with PBS, scraping each well and collecting them in a 1.7 mL Eppendorf tube. The tubes were centrifuged at 14,000 rpm

for 5 min and the supernatant was carefully aspirated. The pellets were either used right away for RNA extraction or frozen at -20°C until needed.

C3H10T1/2 sample preparation for iTRAQ & gel separation: To determine the compounds possible targets, affinity purification was performed. C3H10T1/2 cells were grown in T175 flasks in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin at 37°C/5% CO_2 until confluent. Subsequently, the media was removed and the cells were washed once with PBS before 25 mL/flask DMEM supplemented with 0.5% FBS, 1% penicillin/streptomycin, and 5% Shh-conditioned medium was added. The cells were then incubated for 48 hours at 37°C/5% CO_2. Subsequently, the cells were harvested by washing them once with PBS, trypsinizing and counting. They were centrifuged at 1000 rpm for 5 minutes, the media removed, placed on ice, and the pellets washed once with PBS. The PBS was then removed, and 1 mL lysis buffer (ModRipa with 1% NP-40) per 50 million cells was added to the pellet. The pellets were homogenized by vortexing and incubated on ice for 15 minutes. Subsequently, the mixture was centrifuged at 14,000 rpm for 10 minutes at 4°C. The protein concentration was assessed and 1500 mg (700 mg for gel electrophoresis) protein were brought to 1 mL with lysis buffer and incubated with 6 μL (200x) BRD50837, inactive enantiomer (both 10 mM solution in DMSO), or DMSO for 1 hour on ice. The mixture was then added to a slurry of 45 μL beads with immobilized active compound for each sample and incubated overnight at 4°C under rotation. Subsequently, the beads were centrifuged at 3000 rpm for 5 minutes at 4°C and washed once with 1 mL lysis buffer and twice with 1 mL cold PBS. For each washing step, they were incubated under rotation for 10 minutes and subsequently centrifuged

again. For iTRAQ, the PBS was removed and the beads were stored frozen at -80°C in about 10 μL of PBS. For gel electrophoresis, the PBS was removed and 40 μL of 2X NuPage LDS Sample buffer was added to each sample. The samples were subsequently boiled for 10 min at 95°C and moved forward to gel separation.

iTRAQ sample tagging and mass spectrometry: 90 μL digestion buffer (2M urea, 50 mM Tris HCl) was added to thawed beads. Proteins bound to beads were reduced with 2 μL 500 mM DTT for 30 minutes and alkylated with 4 μL of 500 mM Iodoacetamide for 45 minutes in the dark. 2 μg of sequencing grade trypsin was incubated overnight at room temperature. Mixture was quenched with 20 μL 10% FA, and supernatant and two 50 μL washes with digestion buffer were collected into a fresh tube. Acetonitrile was added to mixture to a final concentration of 3%, and samples were desalted on Oasis cartridges. Concentrated peptide eluents were dried and iTRAQ labeled using Sciex iTRAQ kits (standard manufacturer protocols were followed). Peptides were analyzed for labeling efficiency, labeled peptides were quenched, combined, and dried. Combined peptides were fractionated using strong cation exchange and desalted prior to mass spectrometry analysis. Samples were run on a Q Exactive with a 110 minute UPLC gradient.

Gel electrophoresis and western blotting: the samples were loaded on a Bolt 4-12% Bis-Tris Plus gel and run in MES SDS running buffer at 175V for 30 minutes. Proteins were transferred to nitrocellulose membranes (0.2 μm pore size) using the iBlot system for 7.5 min. After transfer, total protein load was visualized using Ponceau S solution according to manufacturer's protocol. Membranes were destained using a 0.1 M solution of NaOH and washed briefly using TBST (tris-buffered saline and tween 20)

buffer. Membranes were blocked in 5% milk in TBST buffer at 4°C overnight. Subsequently, membranes were washed three times for ten minutes with TBST buffer and incubated with 1° antibody (rabbit IgG, anti-GSTM1) at 4°C overnight. Membranes were again washed three times for ten minutes with TBST buffer and incubated with 2° antibody (anti-rabbit IgG) for one hour at room temperature. Finally, membranes were washed three times for ten minutes with TBST buffer, developed using a Thermo Scientific SuperSignal West Femto Substrate kit and visualized using chemiluminescence settings.

Data Analysis

Dose response curves (C3H10T1/2 and Ptch$^{-/-}$ cells): As described in Chapter 2.

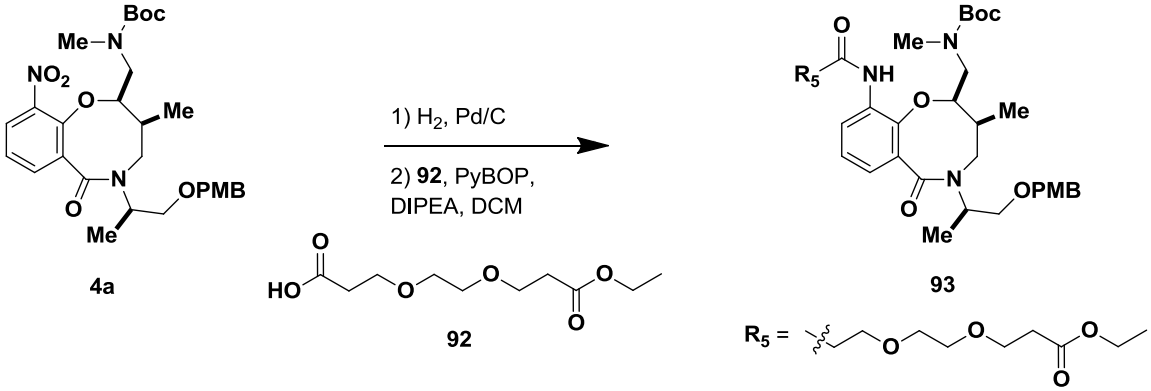
Gene expression data SuFu$^{-/-}$ cells: For each sample, four technical replicates were measured and C_T values for each replicate were calculated using the instrument's analysis feature. C_T values were subsequently averaged for each sample and the fold expression changes normalized within each experiment to SuFu$^{+/+}$ cells was calculated using the $2^{-\Delta\Delta C}{}_T$ method.[19] Experiments were done three times in duplicate on separate days.

iTRAQ: Raw MS files were extracted, searched against a Uniprot mouse database, and validated using Agilent Spectrum Mill MS Proteomics Workbench. iTRAQ mix (N-term, K) was selected as a fixed modification in extraction and MS/MS search. Following export, non-mouse proteins and proteins with one peptide were removed, and remaining proteins were median normalized. A moderated T test was performed to identify significant protein interactors. A protein was deemed significant if its p value

was higher than 0.02 and its average ratio was in the 95th percentile or higher. All proteins and their enrichments are listed in Suppl. Table S4.1 (digital).

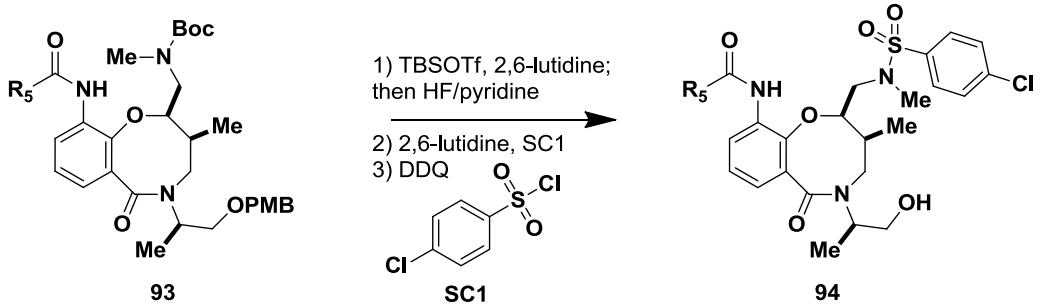
Synthetic Procedures

Protocol analog linker **94**



4a (0.2 g, 1.0 eq) in EtOH (7.2 mL, 0.05 M) was stirred with palladium (10% on activated carbon, 3.8 mg, 0.10 eq) at 35°C under H_2. After the starting material was completely consumed (as monitored by LC-MS, 1-2 h), the mixture was cooled, filtered through celite, and concentrated. The crude material was dissolved in DCM (3.6 mL, 0.1 M) under N_2 and PyBOP (0.28 g, 1.5 eq), and diisopropyl ethylamine (DIPEA) (0.19 mL, 3.0 eq) was added. The mixture was cooled to 0°C before 3-(2-(3-ethoxy-3-oxopropoxy)ethoxy)propanoic acid (**92**) was added. The mixture was allowed to come to room temperature and stirred for 15 h before being quenched with water and extracted with DCM. The combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The crude product was purified by flash chromatography on silica gel (0-100% EtOAc in hexanes) to give the product **93** with 78% yield.

93: ^{1}H NMR (400 MHz, $CDCl_3$, room temperature) δ 8.92 (s, 1H), 8.31 (d, *J* = 7.4 Hz, 1H), 7.22 (d, *J* = 8.0 Hz, 2H), 7.12 (d, *J* = 7.5 Hz, 1H), 6.90 - 6.79 (m, 3H), 5.10 – 4.95 (m, 1H), 4.72 – 4.60 (m, 1H), 4.56 – 4.46 (m, 1H), 4.46 – 4.37 (m, 1H), 4.31 – 4.17 (m, 1H), 4.15 – 4.04 (m, 2H), 3.89 – 3.80 (m, 2H), 3.76 (s, 3H), 3.73 – 3.65 (m, 2H), 3.65 – 3.50 (m, 6H), 3.24 – 3.15 (m, 2H), 2.93 – 2.80 (m, 1H), 2.79 – 2.68 (m, 1H), 2.56 2.48 (t, *J* = 6.1 Hz, 2H), 2.45 (s, 3H), 2.43 – 2.35 (m, 1H), 2.02 – 1.90 (m, 1H), 1.33 (s, 9H), 1.21 (t, *J* = 6.9 Hz, 3H), 1.16 (d, *J* = 6.5 Hz, 3H), 0.90 (d, *J* = 6.2 Hz, 3H).).^{13}C NMR (101 MHz, $CDCl_3$, room temperature) δ 171.76, 170.45, 170.20, 159.41, 156.78, 144.81, 129.90, 129.49, 128.05, 126.33, 122.56, 120.77, 120.63, 113.88, 79.91, 74.15, 72.76, 71.65, 70.29, 70.18, 67.44, 66.64, 60.53, 55.33, 51.36, 49.60, 47.55, 37.80, 35.09, 35.05, 34.36, 28.37, 14.82, 14.22, 10.94. LRMS-ES$^+$ *m/z*: 744.46 (M^+).



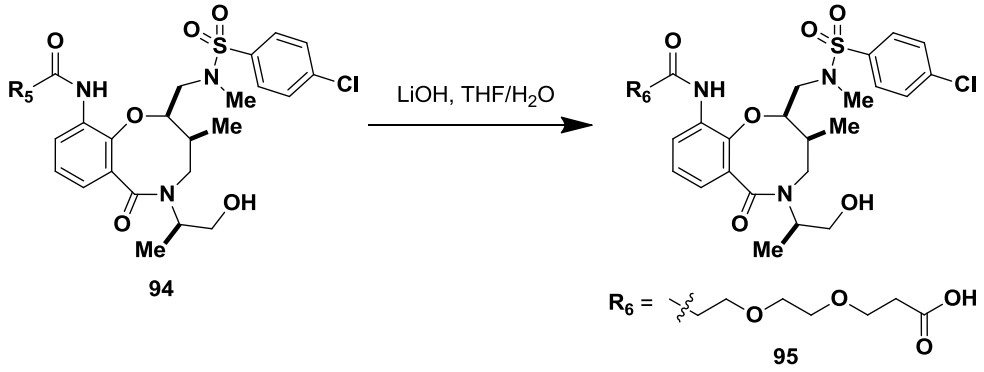
93 (135 mg, 1.0 eq) was dissolved in DCM (1.8 mL, 0.1 M) and 2,6-lutidine (0.09 mL, 4.0 eq) and TBSOTf (0.125 mL, 3.0 eq) were added at room temperature. The reaction was stirred for 2h and subsequently quenched with saturated NH_4Cl solution. The mixture was extracted with EtOAc and the combined organic extracts were dried over $MgSO_4$, filtered, and concentrated to produce the crude silyl carbamate which was dissolved again in THF (1.8 mL, 0.1 M). HF/pyridine (70%, 85 μL, 1.0 eq) was added and the mixture was stirred for 40 min at room temperature. The reaction was then

quenched with saturated NH_4Cl solution and extracted with EtOAc. The combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The crude material was dissolved in DCM (7.3 mL, 0.025 M) and 2,6-lutidine (0.13 mL, 6.0 eq) was added and the mixture was cooled to 0°C. The sulfonyl chloride **SC1** (0.12 g, 3.0 eq) was added in minimal DCM. The reaction was stirred at 0°C overnight and subsequently quenched with water. The mixture was extracted with DCM, dried over $MgSO_4$, filtered, and concentrated. The product was filtered through a short silica plug (0-100 % EtOAc in Hexanes) to remove polar by-products. The product was then dissolved in a 7:1 DCM and pH 7 buffer solution (8 mL, 0.02 M). The mixture was cooled to 0°C and DDQ (0.06 g, 1.5 eq) was added. The reaction was stirred at 0°C for another 10 mins and then brought to room temperature and stirred overnight. The reaction was quenched with water and extracted with DCM. The combined organic extracts were washed with saturated $NaHCO_3$ solution and subsequently, activated carbon was added. The mixture was filtered through celite and the filtrand was washed with hot DCM several times. The filtrate was concentrated and the product was purified by flash chromatography on silica gel (0-100% EtOAc in hexanes) to give the product **94** in 32% yield.

94: ^{1}H NMR (400 MHz, $CDCl_3$, room temperature) δ 8.65 (s, 1H), 8.40 (d, *J* = 7.2 Hz, 1H), 7.69 (d, *J* = 8.2 Hz, 2H), 7.48 (d, *J* = 8.3 Hz, 2H), 7.13 (d, *J* = 7.0 Hz, 1H), 6.93 (t, *J* = 7.9 Hz, 1H), 5.00 – 4.87 (m, 2H), 4.16 – 4.06 (m, 2H), 3.95 – 3.78 (m, 4H), 3.76 – 3.67 (m, 2H), 3.66 – 3.54 (m, 4H), 3.54 – 3.42 (m, 1H), 3.36 – 3.20 (m, 2H), 2.79 – 2.69 (m, 3H), 2.68 (s, 3H), 2.56 – 2.49 (m, 2H), 2.05 – 1.96 (m, 1H), 1.26 – 1.15 (m, 6H), 0.98 (d, *J* = 6.2 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$, room temperature) δ 172.14, 171.67,

170.15, 144.57, 139.46, 136.62, 129.68, 128.61, 128.16, 126.39, 122.80, 121.12, 120.65, 73.03, 70.38, 70.25, 67.30, 66.84, 66.72, 60.56, 52.89, 52.60, 47.70, 38.00, 35.46, 35.28, 35.20, 14.31, 14.17, 11.11. LRMS-ES$^+$ *m/z*: 698.40 (M^+).

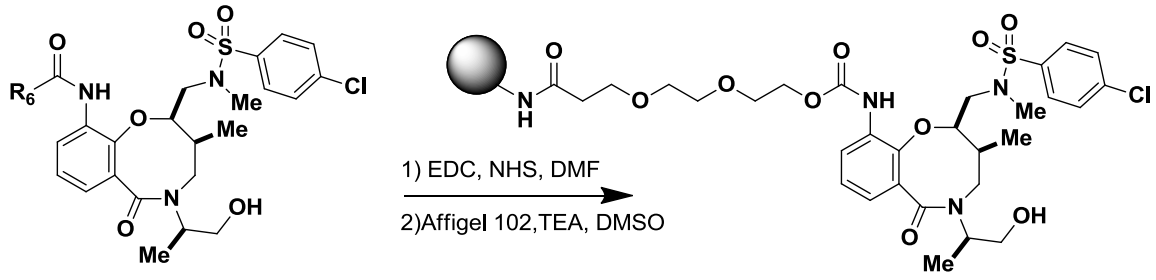
Protocol compound activation and bead loading



94 (18.5 mg, 1.0 eq) was dissolved in a 1:1 mixture of THF/water (800 μL, 0.03 M) and LiOH (3.2 mg, 5.0 eq) was added. The mixture was stirred at room temperature overnight. The mixture was quenched with saturated NaHCO3, and the aqueous phase was acidified with 1M HCl to pH1. The mixture was extracted with EtOAc and combined organic extracts were dried over $MgSO_4$, filtered, and concentrated. The concentrate was purified by flash chromatography on silica gel (0-10% MeOH in DCM) to give the product **95** in 73% yield.

95: ^{1}H NMR (400 MHz, $CDCl_3$, room temperature) δ 8.69 (s, 1H), 8.41 (d, *J* = 7.3 Hz, 1H), 7.72 (d, *J* = 8.1 Hz, 2H), 7.49 (d, *J* = 8.2 Hz, 2H), 7.13 (d, *J* = 7.4 Hz, 1H), 6.94 (t, *J* = 7.8 Hz, 1H), 5.04 – 4.87 (m, 2H), 3.99 – 3.75 (m, 4H), 3.74 – 3.66 (m, 2H), 3.65 – 3.54 (m, 4H), 3.54 – 3.44 (m, 1H), 3.33 – 3.20 (m, 2H), 2.82 – 2.73 (m, 3H), 2.72 (s, 3H), 2.57 – 2.46 (m, 2H), 2.10 – 1.95 (m, 1H), 1.19 (d, *J* = 6.4 Hz, 3H), 0.98 (d, *J* = 6.1 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$, room temperature) δ 174.14, 172.44, 170.47, 144.67,

139.50, 136.55, 129.72, 128.68, 128.07, 126.52, 122.97, 121.10, 120.57, 73.01, 70.42, 70.07, 67.34, 66.73, 66.55, 52.90, 52.73, 47.82, 38.03, 35.42, 35.35, 34.91, 14.19, 11.10. LRMS-ES$^+$ *m/z*: 767.34 (M^+).



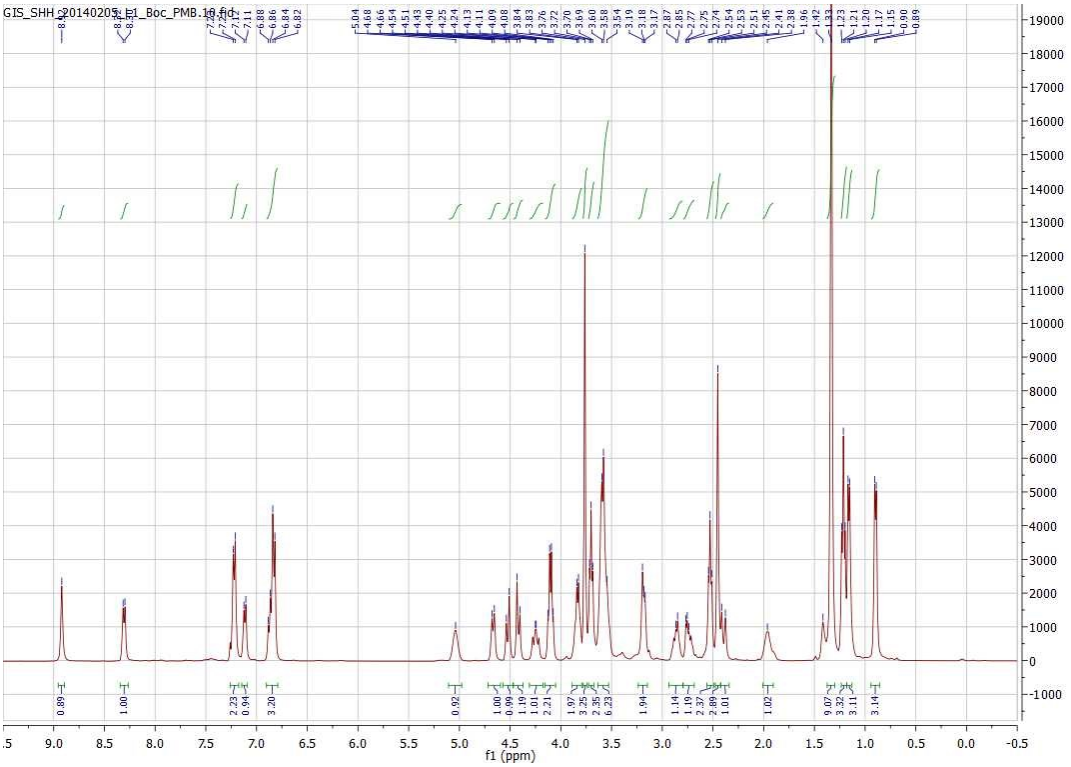
95 **96**

95 (5.9 mg, 1.0 eq) was dissolved in DMF (400 μL, 0.02 M) and *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (EDC, 3.4 mg, 2.0 eq) and N-hydroxy succinimide (NHS, 2.0 mg, 2.0 eq) was added. The reaction was heated to 40°C and stirred for 3 h. After 3 h, the mixture was 90% activated by LCMS and did not react further. The mixture was directly used for the next step.

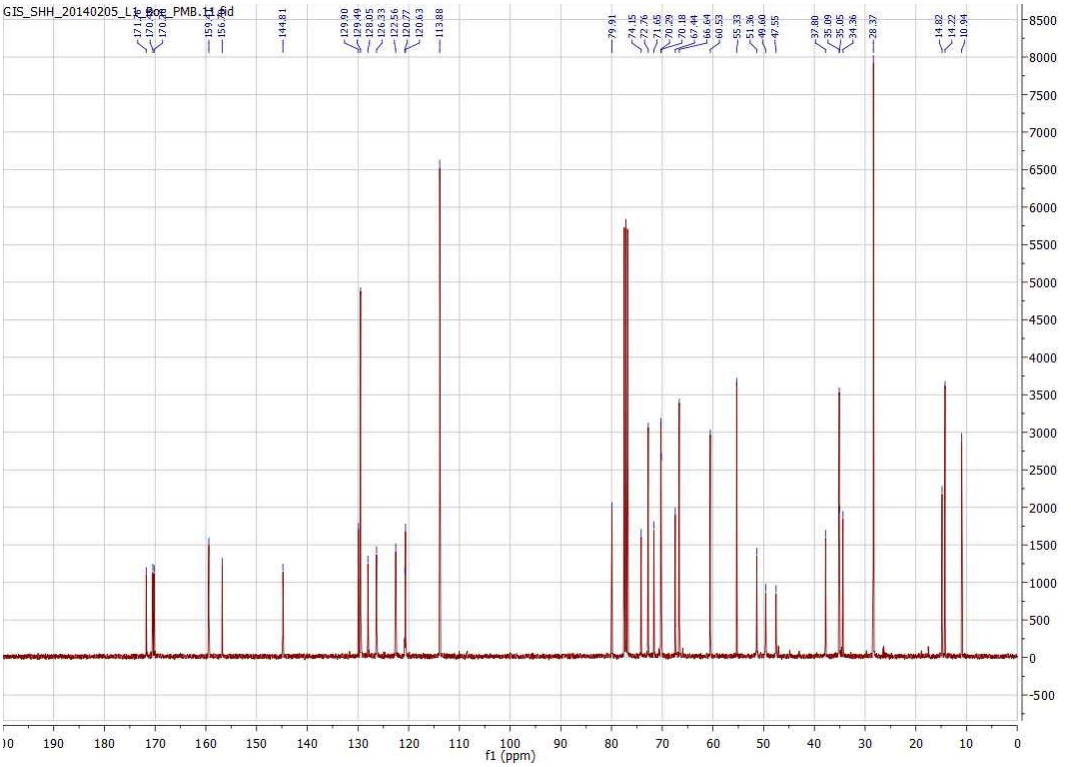
1 vial of Affigel 102 beads (1 mL, 12 μmol) was washed twice with DMSO and three times with anhydrous DMSO. The beads were resuspended in DMSO (500 μL) and activated compound **95** in DMF from above (72 μL, 12.5% loading) was added. Triethylamine (17 μL, 10 eq) was added and the vials were shaken gently at room temperature. After 90 min, no starting material was detectable by LCMS. The beads were washed 4 times with DMSO, resuspended in PBS (500 μL), and stored at 4°C.

Spectra

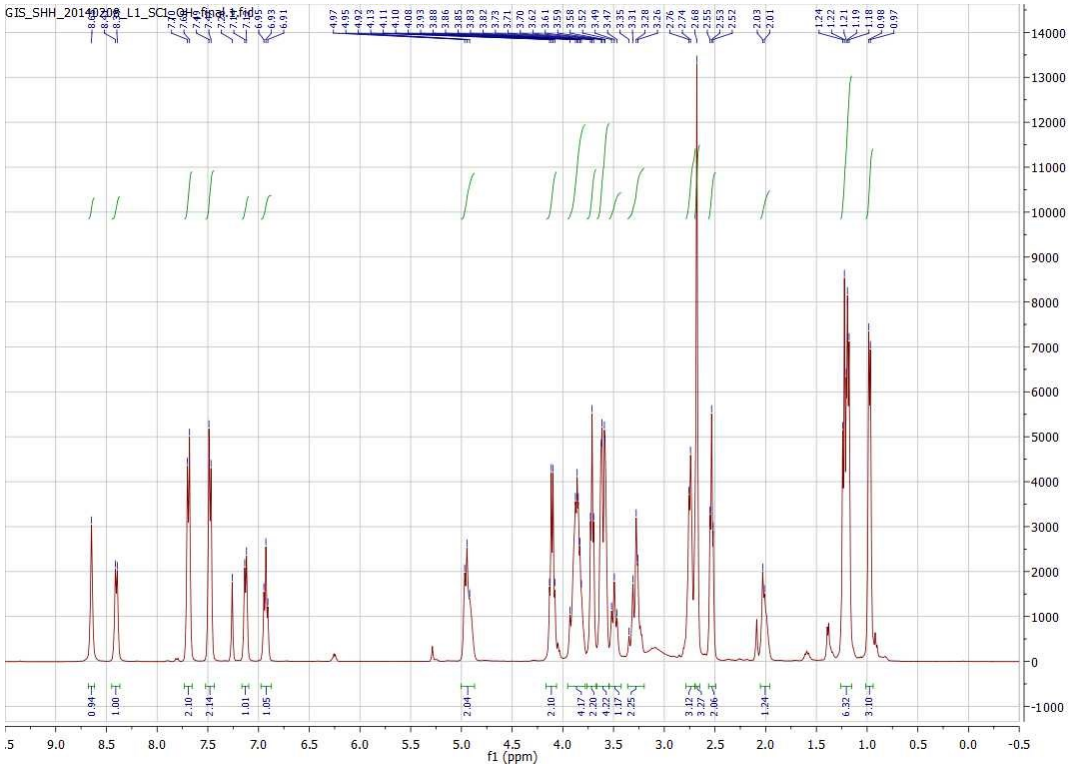
93 ^{1}H NMR:



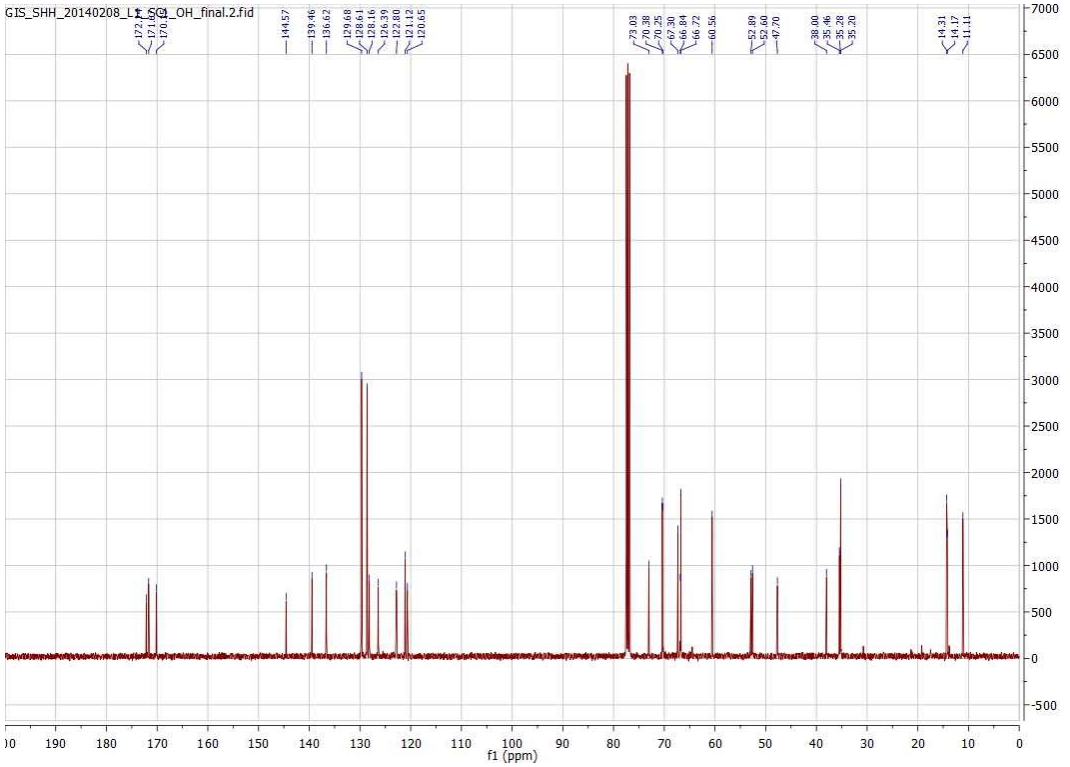
93 ^{13}C NMR:



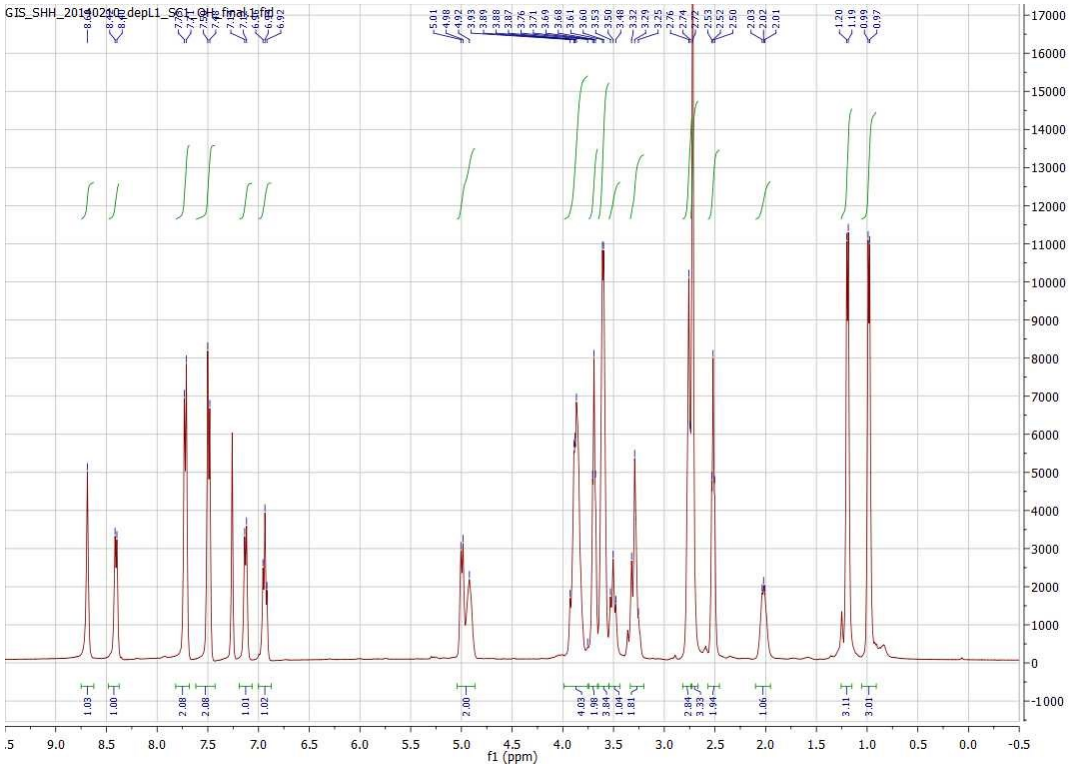
94 ^{1}H NMR:



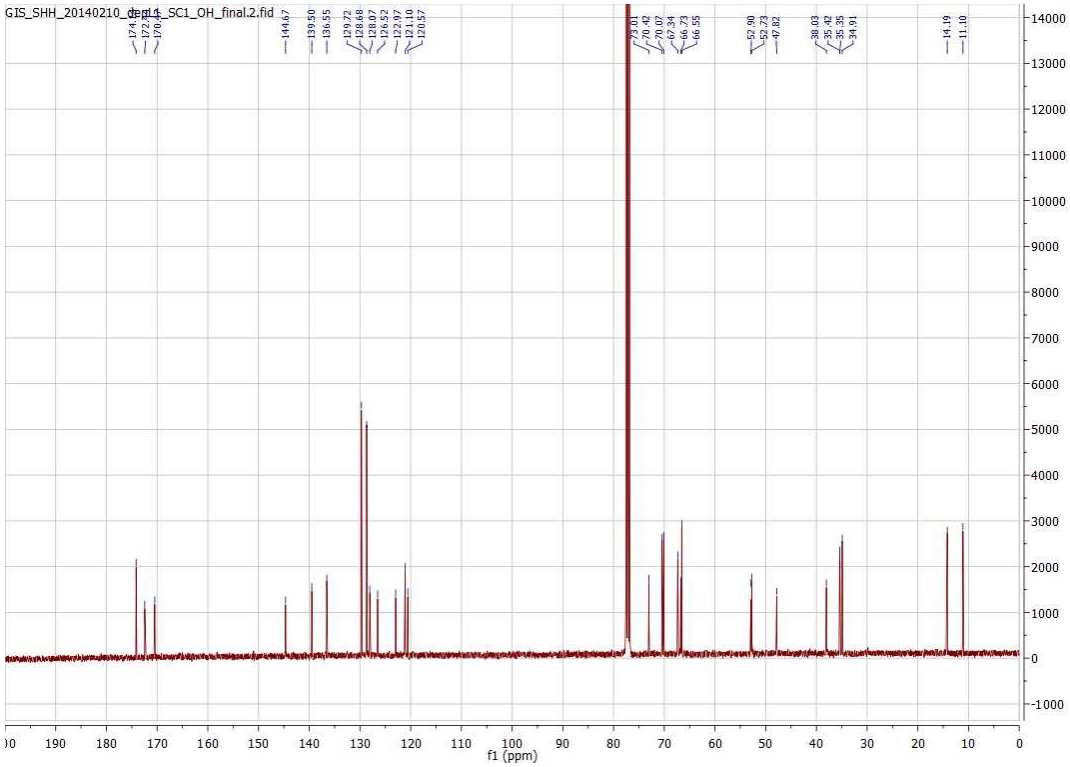
94 ^{13}C NMR:



95 ^{1}H NMR:



95 ^{13}C NMR:



References

(1) Reprinted (adapted) with permission from Schaefer, G. I. *et al.* "Discovery of Small-Molecule Modulators of the Sonic Hedgehog Pathway." *J. Am. Chem. Soc.*, *135*, 9675-9680. Copyright (**2013**) American Chemical Society.

(2) Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. "Small molecule modulation of Smoothened activity." *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14071-14076.

(3) Taipale, J. *et al.* "Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine." *Nature* **2000**, *406*, 1005-1009.

(4) Sinha, S. & Chen, J. K. "Purmorphamine activates the Hedgehog pathway by targeting Smoothened." *Nat. Chem. Biol.* **2006**, *2*, 29-30.

(5) Svärd, J. *et al.* "Genetic elimination of Suppressor of fused reveals an essential repressor function in the mammalian Hedgehog signaling pathway." *Dev. Cell* **2006**, *10*, 187-197.

(6) Lauth, M., Bergström, A., Shimokawa, T. & Toftgård, R. "Inhibition of GLI-mediated transcription and tumor cell growth by small-molecule antagonists." *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8455-8460.

(7) Yauch, R. L. *et al.* "A paracrine requirement for hedgehog signalling in cancer." *Nature* **2008**, *455*, 406-410.

(8) Zhang, X. *et al.* "Cyclopamine inhibition of human breast cancer cell growth independent of Smoothened (Smo)." *Breast Cancer Res. Treat.* **2009**, *115*, 505-521.

(9) Robarge, K. D. *et al.* "GDC-0449-a potent inhibitor of the hedgehog pathway." *Bioorg Med Chem Lett* **2009**, *19*, 5576-5581.

(10) Schenone, M., Dancik, V., Wagner, B. K. & Clemons, P. A. "Target identification and mechanism of action in chemical biology and drug discovery." *Nat. Chem. Biol.* **2013**, *9*, 232-240.

(11) Ziegler, S., Pries, V., Hedberg, C. & Waldmann, H. "Target identification for small bioactive molecules: finding the needle in the haystack." *Angew. Chem., Int. Ed.* **2013**, *52*, 2744-2792.

(12) Ross, P. L. *et al.* "Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents." *Mol. Cell. Proteomics* **2004**, *3*, 1154-1169.

(13) Monica Schenone & Hartmann, E., verbal communication, based on 5-10 iTRAQ experiments in different systems, (personal communication to Giannina Schaefer, 04/06/2014).

(14) Townsend, D. M. & Tew, K. D. "The role of glutathione-S-transferase in anti-cancer drug resistance." *Oncogene* **2003**, *22*, 7369-7375.

(15) Oakley, A. "Glutathione transferases: a structural perspective." *Drug Metab. Rev.* **2011**, *43*, 138-151.

(16) Sheehan, D., Meade, G., Foley, V. M. & Dowd, C. A. "Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily." *Biochem. J.* **2001**, *360*, 1-16.

(17) Bassilana, F. *et al.* "Target identification for a Hedgehog pathway inhibitor reveals the receptor GPR39." *Nat. Chem. Biol.* **2014**, *10*, 343-349.

(18) Frei, A. P. *et al.* "Direct identification of ligand-receptor interactions on living cells and tissues." *Nat. Biotechnol.* **2012**, *30*, 997-1001.

(19) Livak, K. J. & Schmittgen, T. D. "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." *Methods* **2001**, *25*, 402-408.

Chapter Five

Cancer cell-line profiling to identify cancer genetic and lineage dependencies targeted by small molecules

I mined the literature to decide which compounds could be part of the informer set together with Dr. Jordi Barretina, Dr. Marcin Imielinski, Dr. Stefan Kubicek, Dr. Ben Munoz, Josh Paulk, Dr. Gopal Ramachandran, Dr. Arvind Ramanathan, Dr. Aly Shamji, Dr. Andrew Stern, Dr. Bridget Wagner, and Dr. Stuart Schreiber. Dr. Jaime Cheah, Edmund Price, Dr. Ke Liu, Richard Ebright, Dr. Daisuke Ito, Stephanie Wang, Dr. Abigail Bracha, and I carried out the assay development and profiling runs of the CCLs. Dr. Amrita Basu, Nicole Bodycombe, Dr. Mathias Wawer, Dr. Josh Gilbert, and Dr. Paul Clemons processed and analyzed the data. Dr. Andrew Wilson and Dr. Dineo Khabele confirmed the response of ML210 and ML162. Ted Liefeld implemented the CTRP portal.

In part adapted from Basu, A.; Bodycombe, N.E.; Cheah, J.H.; Price, E.V.; Liu, K. Schaefer, G.I.; Ebright, R.Y.; Stewart, M.L.; Ito, D.; Wang, S.; Bracha, A.L.; Liefeld, T.; Wawer, M.; Gilbert, J.C.; Wilson, A.J.; Stransky, N.; Kryukov, G.V.; Dancik, V.; Barretina, J.; Garraway, L.A.; Hon, C.S-Y.; Munoz, B.; Bittker, J.A.; Stockwell, B.R.; Khabele, D.; Stern, A.M.; Clemons, P.A.; Shamji, A.F.; Schreiber, S.L. “An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules**”**, *Cell*, **2013**, *154*, 1151–1161.[1]

Profiling cancer cell lines against small molecules

In contrast to the directed approach of finding small-molecule probes for a pre-defined pathway presented in the previous chapters, unbiased approaches can uncover novel hypotheses about all areas of cellular biology. Cancer cell-line (CCL) profiling has been employed previously to determine patterns of cancer dependencies. In order to identify cancer dependencies and small molecules that can target these dependencies directly, we profiled the sensitivity of 242 CCLs against an Informer Set consisting of 354 small molecules with annotated targets and activities. The results are provided in the Cancer Therapeutics Response Portal (CTRP; http://www.broadinstitute.org/ctrp) which enables researchers to analyze relationships between features of the CCLs and small-molecule sensitivity.

In order to profile the CCLs' sensitivity to small molecules, we first assembled an Informer Set of compounds. These molecules targeted several distinct cellular proteins and/or had high selectivity for their annotated targets (*e.g.*, rapamycin targeting mTOR[2]). In order to validate possible correlations between genetic features and sensitivity, compounds having different structures but the same target protein (*e.g.*, cyclosporin A and tacrolimus targeting calcineurin[3]), and compounds targeting the same family of proteins, but with differential selectivity (*e.g.*, histone deacetylase inhibitors[4]) were included. If several compounds with similar target profiles were available, priority was given to those in clinical development, strong selectivity data, or with pharmacokinetic data, so that drug development based on the discovered dependencies could be accelerated. The Informer Set consisted of 36 FDA-approved drugs, 54 clinical candidates, and 264 probes (Suppl. Table S5.1 (digital)).

The Cancer Cell Line Encyclopedia collection comprises ~1,000 genetically characterized CCLs. This characterization includes gene expression, amplifications/deletions, somatic mutations in >1,600 cancer genes, and lineage/histological subtypes and is freely available (http://www.broadinstitute.org/ccle). We chose subsets of lineages that overlap with The Cancer Genome Atlas and in published genome-wide RNAi screens.[5] We plated each CCL in its established media at a density optimized during assay development (Suppl. Table S5.2j-l (digital)) and treated with compounds for 72h at eight concentrations. CellTiter-Glo was used to measure cellular ATP levels as a surrogate for viability. The area under percent-viability curves (AUC), which is a function of both relative potency and total level of inhibition for a compound across CCLs, was computed as a metric of sensitivity (Suppl. Table S5.2 (digital), Suppl. Figure S5.1).

To analyze the sensitivity data, we first looked at the distributions of AUCs across all compounds to determine if trends among subpopulations were present (Figure 5.1a). Most CCLs show a differential spread of responses across the Informer Set. However, CCLs within the hematopoietic and lymphoid lineage and suspension CCLs were often more sensitive to compounds tested than other CCLs (Figure 5.1b). To control for potential confounding factors arising from this observation, the analyses of AUC distributions was divided into sets that included all CCLs, as well as sets that excluded specific context-dependent subsets of CCLs (Figure 5.1c, Suppl. Figure S5.1).

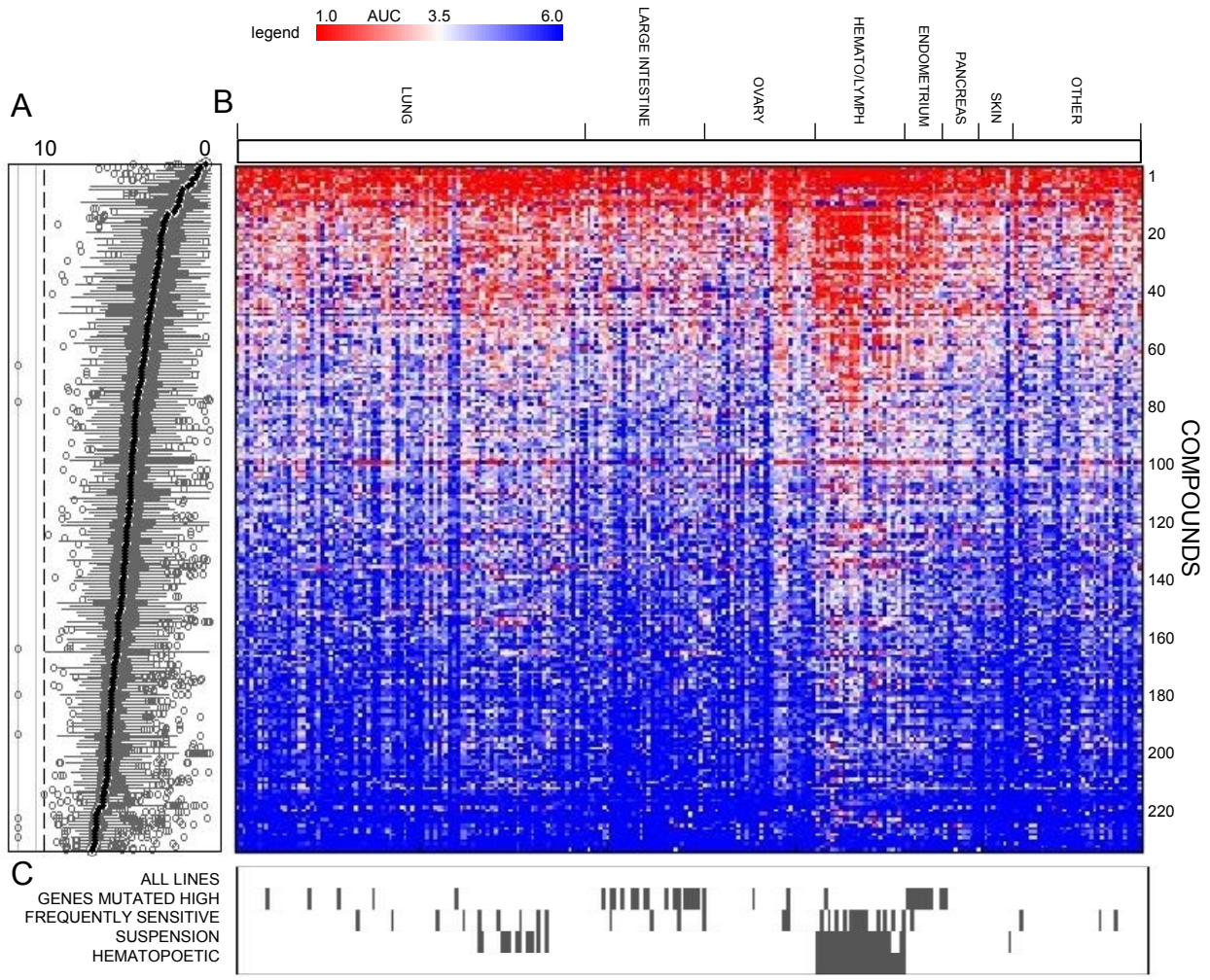


Figure 5.1. Response of CCLs to Informer Set. Sensitivity of 242 CCLs to small-molecule probes/drugs was assessed at dose (CellTiter-Glo) and areas under the concentration-response curve (AUC) were computed. Data are shown as box plots indicating distributions of AUC values for each compound (A) and a heatmap of AUC values (scale represents AUC values ranging between 1 (sensitive; red) and 6 (unresponsive; blue)) (B) for single CCLs (columns) treated with single compounds (rows). Missing numerical values in heatmap were imputed *using a k-nearest neighbors approach*. AUC distributions were analyzed by incorporating context-dependent exclusions (C) of cell lines (grey bars represent excluded cell lines).

For each compound across all CCLs and relevant subsets, statistics-based enrichment analyses were performed. These analyses combined rank-based and parametric tests to identify genetic alterations and cellular features that are significantly enriched among sensitive or unresponsive CCLs. All correlations are available as a

table (Suppl. Table S5.2 (digital)) and those with statistical significance can also be visualized online in the CTRP.

Cancer cell-line profiling identifies known cancer dependencies

In order to properly predict novel, clinically relevant cancer dependencies using CCL profiling, CCLs have to be a good model of tumor responses. This can only be confirmed using patient-derived data under the same treatments. Several known relationships between mutations and sensitivity were identified during our effort. Cell lines with a *BRAF* V600E mutation showed increased sensitivity to P-0850, an analog of the FDA-approved BRAF-V600E inhibitor vemurafenib (Figure 5.2a).[6] Additionally, upon inspection of the unresponsive V600E CCLs, the unresponsiveness could be attributed to previously identified mechanisms of resistance to vemurafenib: 1) enhanced EGFR signaling has been reported to lead to resistance in vemurafenib-treated colon cancers,[7] and the CCL SKMEL28 contains an activating mutation in EGFR; and 2) RKO, a colorectal CCL, has been shown to circumvent dependence on BRAF by producing high levels of hepatocyte growth factor (HGF), which activates CRAF via MET in an autocrine fashion. It has been shown that blocking HGF signaling using a MET inhibitor can resensitize the cells to BRAF-V600E inhibition.[8,9] Thus, studying nonresponsive outliers of other compounds can offer insight into resistance mechanisms.

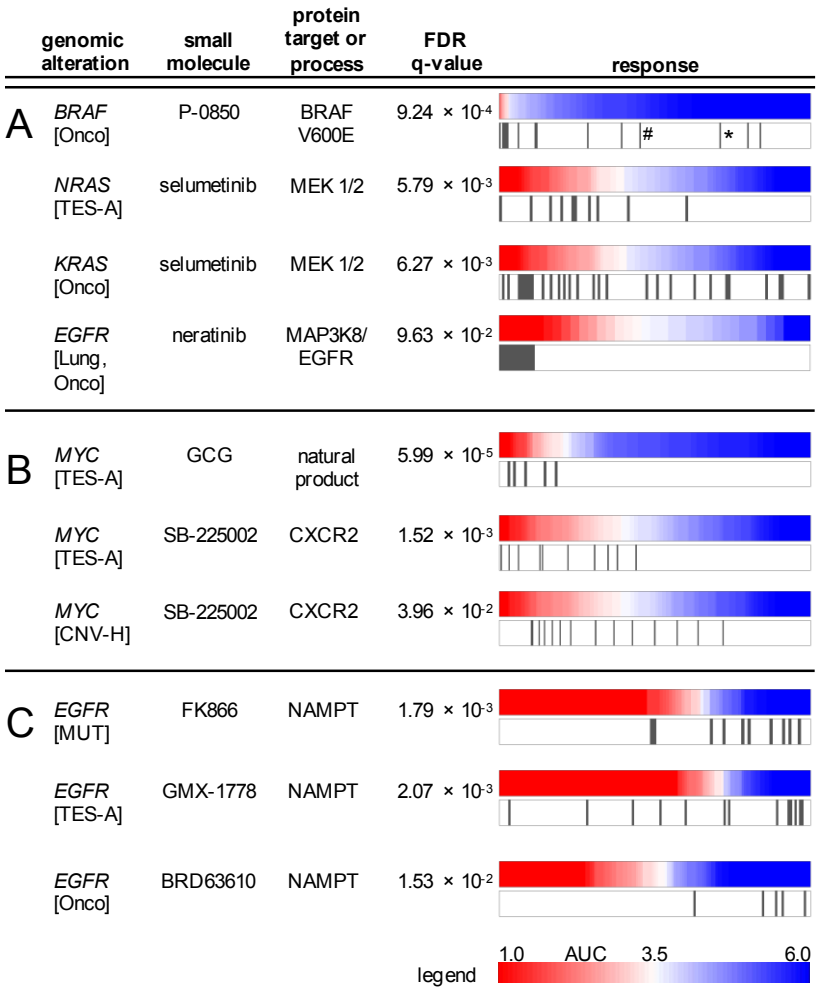


Figure 5.2. Genetic dependencies targeted by small molecules. The distribution of CCL response (AUC values) to compound treatment is represented as a heatmap denoting sensitivity (red) or unresponsiveness (blue) aligned with genomic alterations for corresponding CCLs (gray bars). The CCL profiling identified known clinically drug-targeted genetic dependencies (A) and known drug-resistance mechanisms (BRAF V600E outlier cell lines: *RKO; #SKMEL28). It also suggests dependencies with both mutation and copy number variation in *MYC* (B). Global analysis of the profiled data showed *EGFR*-mutated CCLs are unresponsive to NAMPT inhibitors (C). CNV-H: high-copy number (≥8 copies), TES: all targeted-exome sequencing mutant calls, TES-A: targeted-exome sequencing, non-neutral missense mutations; Onco: Oncomap mutant calls, MUT: any mutation call.

Another known dependence is that of NRAS- and KRAS-mutant CCLs to selumetinib, a MEK1/2 inhibitor. This compound has shown preliminary moderate activity in *KRAS*-mutant patients during clinical trials (Figure 5.2a).[10-12] However, several

mutant CCLs are not sensitive to selumetinib, which suggests that factors other than KRAS/NRAS might also influence the response. As with P-0850, analysis of unresponsive outliers could lead to identification of other biomarkers for selumetinib response.

When looking at genetic features within specific lineages, more correlations were found. For example, neratinib, a dual ERBB2/EGFR inhibitor currently in clinical trials for advanced non-small cell lung cancer,[13] shows increased potency in EGFR-mutant lung CCLs (Figure 5.2a). Therefore, our CCL profiling effort can identify known dependencies, but also detect already-known resistance mechanisms. It is thus reasonable to expect to discover new insights as well.

Novel hypotheses generated by CCL profiling

Several oncogenes are known that have been shown to drive cancers but do not yet have targeted therapies. The profiling effort suggests small molecules that could be used to exploit the dependencies created by these oncogenes. For example, (-)-gallocatechin-3-monogallate (GCG) shows a strong response in CCLs with *MYC* mutations, including those interfering with MYC protein degradation (Figure 5.2b).[14] A GCG analog, epigallocatechin-3-monogallate, has been shown to decrease MYC expression in digestive tract-derive CCLs and mouse tumor models.[15,16] Additionally, mutations in *MYC* and amplifications of *MYC* show sensitivity to an inhibitor of the chemokine receptor CXCR2, SB-225002 (Figure 5.2b). This receptor is implicated in promoting oncogene-induced senescence.[17] Its relationship to MYC biology is not yet

well understood, but the correlation of two kinds of genomic alterations to SB-225002 suggests that further investigations into this connection might be warranted.

While the discovery of novel cancer dependencies with sensitivities to small molecules is of high interest, attention should also be paid to those molecules where mutated genes correlate with unresponsiveness. Several genes showed unresponsiveness to many compounds. For example, *EGFR*-mutated CCLs correlated with unresponsiveness to different inhibitors of NAMPT (Figure 5.2c). Additionally, in a murine model of *KRAS*-mutant lung cancer, *STK11* has been implicated in resistance to docetaxel.[18] Similarly, patients with *BRAF*-mutant metastatic colon cancers show little response to EGFR-targeted therapy.[19] CCL profiling shows that many cell lines with mutations in these genes are unresponsive to a wide variety of compounds. These results suggest that mutations, such as those listed above, might cause resistance to a wide range of small molecules, and could lead to a more guided approach when treating patients with mutations in these genes.

Lineage dependencies of CCLs

Apart from genetic features, lineage features also correlated with sensitivity to different small molecules. For example, ovarian CCLs were very sensitive to ML210 and RSL3, two probes that selectively kill engineered *HRAS*-mutant CCLs (Figure 5.3a). These probes were originally found to selectively kill transformed cell lines BJeLR (HRasG12V, SV40 large T and small T antigens) and DRD (HRasG12V, hTERT, SV40 small T oncoprotein, dominant negative p53, cyclin D1, and mutant CDK4), relative to untransformed controls.[20] However, mutations in *HRAS* did not correlate with sensitivity

to the compounds. The potency of ML210 against five ovarian CCLs was confirmed, using sulforhodamine B to measure cellular protein content as a surrogate for cytotoxicity (Figure 5.3b).[21] This confirmation included three previously untested CCLs.

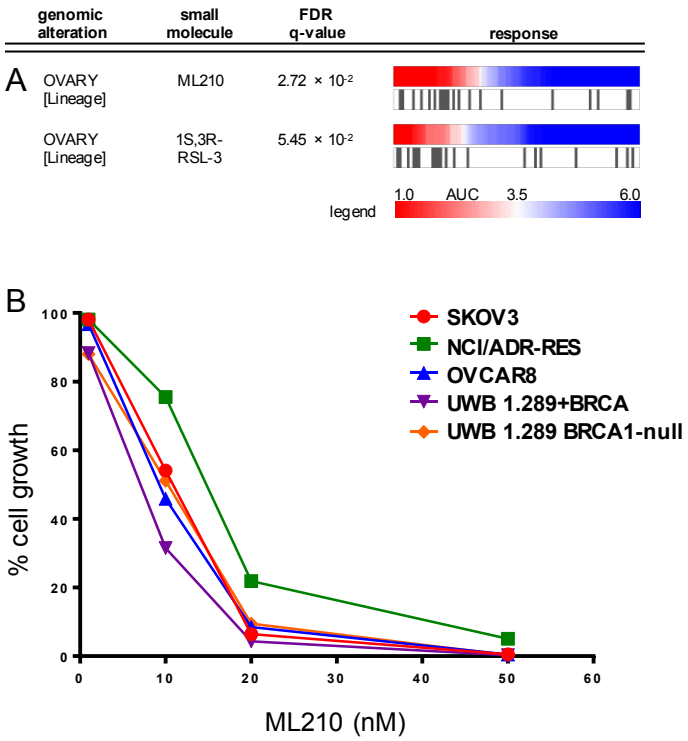


Figure 5.3. Lineage dependencies targeted by small molecules. (a) Ovarian CCLs are highly sensitive to ML210 and RSL3. (b) An expanded panel of ovarian CCLs showed sensivity to ML210 (IC_{50} ~10 nM) independent of the BRCA1 status of the CCLs. All values are shown and generated from three independent experiments run in sextuplicate (values are percent-viability score (Experimental Section)).

RSL3 and related compounds kill cells via a mechanism called ferroptosis.[22] They do, however, also appear to promote markers of apoptosis within this context. CCLs that are sensitive to these compounds might possibly have features also present in the engineered cells described above, which cause this sensitivity to ferroptosis modulators. However, in order to test this hypothesis and find these features, more CCLs will need to be profiled. Multi-feature correlation analyses may also aid in determining features that drive sensitivity.

The same oncogene often creates different dependencies in different lineage contexts (*e.g.*, *BRAF* in melanoma vs. colorectal) and the CTRP allows us to identify dependencies in all CCLs as well as single lineages. One example is *KRAS*-mutant CCLs, which show significant correlation with navitoclax in colorectal, but not all lineages (Suppl. Table S5.2 (digital)). It has been shown that navitoclax synergizes with selumetinib to kill *KRAS*-mutant CCLs, but while some effect can be seen in all lineages, the synergy is most pronounced in colorectal CCLs.[23] In this study, the authors concluded that selumetinib alone in *KRAS*-mutant CCLs is largely cytostatic rather than cytotoxic, explaining the lack of efficacy. Navitoclax however, activates apoptosis and induces cell death in combination with selumetinib. In the assay we conducted CCLs were treated for 72h and viability was measured in ATP levels, a surrogate for cell growth or proliferation. Therefore, relationships between sensitivity and genes involving cytostasis may be identified. While they are very sensitive to selumetinib, *KRAS*-mutant CCLs in our study often only show partial inhibition of ATP levels, a sign of cell growth inhibition. Thus, combining specific cancer genotype-selective cytostatic compounds with other molecules might result in greater efficacy. Corcoran *et al.* describe a screening approach for detecting such combinations. This approach also demonstrates that the level of inhibition is important in analyses of existing data and motivates future incorporation of assays for cell death into data collection.[23]

While our approach mostly associates sensitivity to small molecules with single features, multiple features can be influencing a CCLs response to a compound. This fact can cause challenges in interpreting whether a factor is causal. For instance, many hematopoietic/lymphoid CCLs show more sensitivity to many compounds than other

lineages. One example is *MYC*-mutant CCLs which are very sensitive to the BRD4 inhibitor JQ-1, but many of which are also hematopoietic/lymphoid CCLs. It is unclear whether the genetic or the lineage feature determines sensitivity more. A larger set of mutant CCLs will be necessary to study the mutations separately within hematopoietic/lymphoid and solid tumor CCLs. In general, the CTRP allows for this kind of analysis. One example is SB-225002, which shows good potency in *MYC*-amplified CCLs, no matter if all lineages or only those from solid tumors are analyzed.

Global clustering of compounds

As we had included structurally different compounds with similar mechanisms of action, we wanted to see in a global analysis what insights about dependencies we could gather studying connections between genes and sets of compounds. To ensure that these clustering experiments were not dominated by relatively weaker connections, we limited the analyses to those with false discovery rates (FDR) of $q<0.025$. We used hierarchical clustering of compounds based on connections to genetic features profiles (Suppl Table S5.3 (digital)). Several compounds with similar mechanism of action clustered together, such as: PRIMA-1 and PRIMA-1-Met (re-activators of mutant p53 signaling), FK866 and GMX-1778 (NAMPT inhibitors), neopeltolide and leucascandrolide A (modulators of respiration), and teniposide and etoposide (topoisomerase inhibitors) (Suppl. Figure S5.2). The fact that small molecules with similar mechanism of action result in similar CCL responses also shows that the data is consistent and repeatable.

Other CCL profiling efforts

CCLs have been used as model systems for human cancer for several decades, but CCL profiling is a more recent development. Several large-scale efforts have been implemented, often with different selection of CCLs, growth conditions, data collection (*e.g.*, readouts), data filtering (*e.g.*, for potentially confounding CCLs) and analysis, and formulation of questions in controlled computational experiments. This may lead to different results and interpretations. It is encouraging that using our approach, known connections (*e.g.*, *KRAS*-*NRAS*/selumetinib, *BRAF*/V600E inhibitors) can be identified. While not all correlations with sensitivity overlap between ours and others' studies, there is a significant overlap in the results. For example, Garnett *et al.* also identified novel connections between GW-843682X, a PLK1 inhibitor and *CDKN2A* (sensitivity) and nutlin-3 and *TP53* (unresponsiveness),[24] showing that results are reproducible. Connections that differ between studies (*e.g.*, NOTCH1/navitoclax (Garnett *et al.*) vs CTNNB1/navitoclax (CTRP)) demonstrate that care should be used when interpreting the data, putting emphasis on interpretation within the given parameters and confirming conclusions with well-designed follow-up studies.

Conclusion

CCL profiling enables us to identify novel cancer dependencies, small molecules to target them, and insights into unresponsiveness or resistance to certain compounds. Known dependencies were detected, and many other can be discovered. This profiling effort employed an Informer Set of small molecules that not only targeted mutations directly, but also non-altered proteins that CCLs have developed dependencies on. This

approach of targeting cancer differs from a related targeting of 'non-oncogene addictions', or "hallmarks" of cancers,[25] which does not pay attention to specific genomic alterations.[26]

With more CCLs and compounds being profiled, the predictions that can be made from the data will grow stronger. New probes and drugs, as well as combinations of compounds, across a larger set of CCLs will expand the number of hypotheses generated. Additionally, new analyses of the existing data can also uncover novel hypotheses, especially when including other cellular features such as gene expression, metabolic, proteomic, and epigenetic profiles.

Experimental Section

Materials and Methods

Cell Lines and Reagents: cells were purchased from the Broad Institute Biological Samples Platform or ATCC (Manassas, VA, USA). All CCLs were grown at 37°C/5% CO_2 in their specified medium (Suppl. Table S5.2 (digital)). All CCLs were tested for mycoplasma using a Takara PCR Mycoplasma Detection Kit (Otsu, Japan). Media components and supplements were purchased from ATCC (VA, USA), Cellgro (VA, USA), Life Technologies (Carlsbad, CA, USA), EMD Millipore Chemicals (Billerica, MA, USA), or Invitrogen (Carlsbad, CA, USA).

Assay Treatments and Readouts: 354 small molecules were assembled for the Informer Set after careful evaluation of probe-development literature (seminars, journals, NIH Molecular Libraries Initiative Probe Reports, and patents). These molecules perturb targets and processes on which cancer cells may become dependent. They were

purchased from Active Biochem (Maplewook, NJ, USA), AnalytiCon (Potsdam, Germany), Asinex (Winston-Salem, NC, USA), Avanti (Alabaster, AL, USA), Axon MedChem (Reston, VA, USA), BioVision (Milpitas, CA, USA), Cayman Chemicals (Ann Arbor, MI, USA), Cellagen (San Diego, CA, USA), ChemBridge (San Diego, CA, USA), ChemDiv (San Diego, CA, USA), Chemietek (Indianapolis, IN, USA), EMD Millipore Chemicals (Billerica, MA, USA), Enamine (Monmouth Jct, NJ, USA), Enzo Lifesciences (Farmingdale, NY, USA), Fermentek (Jerusalem, Israel), InterBioScreen (Moscow, Russia), LC Laboratories (Woburn, MA, USA), Maybridge (Trevillet, UK), Medkoo (Chapel Hill, NC, USA), MicroSource (Gaylordsville, CT, USA), Princeton Biomolecular Research (Princeton, NJ, USA), Selleck (Houston, TX, USA), Sigma-Aldrich (St Louis, MO, USA), Specs (Delft, Netherlands), Symansis (Timaru, New Zealand), Tocris (Bristol, UK), Toronto Research Chemicals (North York, ON, Canada), Vitas-M (Moscow, Russia), WuXi AppTec (Shanghai, China), synthesized by Broad internal chemists, or provided directly by external collaborators. A list of all compounds, with annotated targets and structures, is provided (Suppl. Table S5.1 (digital)). CellTiter-Glo reagent was purchased from Promega (Madison, WI, USA). Plate Reader instruments and settings used were: luminscence settings (384-well): Perkin Elmer (Waltham, MA, USA) Envision 2104 Multilabel Plate Reader, luminescence mirror, luminescence 700 emission filter, 0.1 s measurement time, 6.5 mm measurement height. Absorbance (510 nm) settings: Molecular Devices Spectramax M5 spectrophotometer (Sunnyvale, CA, USA), absorbance wavelength: 510 nm. Compounds were pin-transferred using a CyBio (Jena, Germany) Cy-Bi Well Vario.

Assay Protocols

Assay development: To determine the optimal plating density for profiling, treatment of each CCL with staurosporine at various densities was measured using CellTiter-Glo. CCLs were plated at 500-2,000 cells/well (adherent) or 500-5,000 cells/well (suspension) in opaque, white 384 well plates at 30 μL/well in their specified medium (Suppl. Table S5.2 (digital)). Cells were incubated for 24 h at 37°C/5% CO_2 and pinned with various concentrations of 100 nL/well staurosporine (16-pt, 1.67-fold dilution series, 16 replicates/concentration). The plates were subsequently incubated at 37°C/ 5% CO_2 for 72 hours. After the incubation period was over, the plates were removed from the incubator and allowed to come to room temperature. Subsequently, 30 μL of a 1:2 dilution of CellTiter-Glo reagent in PBS was added to each well and the plates were shaken at room temperature for a few seconds. The plates were then incubated at room temperature for at least 10 min and read out using luminescence settings (384-well).

Cancer Cell-Line (CCL) Profiling: CCLs were plated at the density optimized during assay development in opaque, white 384 well plates at 30 μL/well in their specified medium (Suppl. Table S5.2 (digital)). Cells were incubated for 24 h at 37°C/ 5% CO_2 and pinned with compound plates (8-pt, 2-fold concentration ranges as defined by literature review, each compound plate was pinned against two assay plates) at 100 nL/well. The plates were subsequently incubated at 37°C/5% CO_2 for 72 hours. After the incubation period was over, the plates were removed from the incubator and allowed to come to room temperature. Subsequently, 30 μL of a 1:2 dilution of CellTiter-Glo reagent in PBS was added to each well and the plates were shaken at room

temperature for a few seconds. The plates were incubated at room temperature for at least 10 min and read out using luminescence settings (384-well).

Confirming sensitivity of ovarian CCLs to ML210: Five ovarian CCLs, SKOV3, OVCAR8, NCI/ADR-RES, UWB1.289 (*BRCA1* null), and UWB1.289+BRCA (stably expressed *BRCA1*-WT) CCLs were plated at 2,000 cells/well in opaque, white 384 well plates in their specified medium (Suppl. Table S5.2 (digital)). Cells were incubated for 24 h at 37°C/5% CO_2 and pinned with compound plates (four concentrations of ML210) at 100 nL/well. The plates were subsequently incubated at 37°C/5% CO_2 for 72 hours. After the incubation period was over, cell viability was assayed using a sulforhodamine B assay as previously described[21] with minor modifications for seeding at the chosen density. The plates were read out using absorbance (510 nm) settings. Three independent runs were performed with six replicates at each compound dose for each run.

Data Analysis

Assay development: Mock treatments for each plate were averaged and used to calculate a percent response for each staurosporine treatment concentration on the corresponding plate. Percent responses were averaged for each concentration point at each cellular density and standard deviation was measured. Z' factors were calculated at each concentration point and compared between the cellular densities. Densities with the largest dynamic detection window (percent inhibition minimized when Z' = 0; if several densities were similar, lower confluency was the deciding factor) were taken as optimized seeding density.

AUCs were used to calculate sensitivity summary scores (as described below). Variability in the AUC potentially stemming from using different cell densities across CCLs was calculated to be of the same magnitude as day-to-day variability in AUCs. Thus, performing the assays at densities with robust signal detection was favored.

Data processing: A compound-performance score (D-score) was computed for each concentration of compound.[27] This score expresses effect size as a weighted average of differences between control and treatment, and, by estimating how likely it is that an observed effect size is different from effects expected for mock-treatment (DMSO) in the assay, statistical significance. Using any number of replicates of a compound treatment across plates and days, the method of maximum likelihood combines these into the weighted average and its uncertainty. The D-score is the ratio of difference to the uncertainty, a normalized value for each compound in an assay. This weighted average was computed on log-transformed small-molecule sensitivity data and an appropriately weighted average of ratios (i.e., weighted fold-change) of compound-treated to mock-treated wells was obtained. After re-exponentiation, this number was used as percent-viability score. Dose-response curves of percent-viability scores were fit using the MATLAB curve-fitting tool box (smooth cubic splines for multivariate data). Areas-under-curve (AUCs) for each compound were used as a measure of sensitivity in subsequent enrichment and regression analyses.

Genetic data: Annotations of CCLs are publicly available (http://www.broadinstitute.org/ccle). This includes gene expression (Affymetrix GeneChip Human Genome U133 Plus 2.0 Array), copy number (Affymetrix Genome-Wide Human SNP Array 6.0), and mutation status from hybrid capture followed by

massively parallel sequencing of >1,600 genes and from mass spectrometric genotyping (OncoMap) for 492 mutations in 33 oncogenes/tumor suppressors.[28] Genetic diversity was illustrated by analyzing frequency distributions of number of genes across all CCLs and the number of unique lesions for each gene (Suppl. Figure S5.1)

Enrichment analysis: p-values that quantify the enrichment of genetic alterations relative to ranked sensitivities measured for a single compound across many CCLs were calculated using a sorting-based enrichment-scoring algorithm.[29] The p-value for each compound and each genetic feature corresponds to the likelihood of seeing that pattern of alterations (or a stronger one) enriched among the ranked sensitivities by chance. However, these scores do not take the relative or absolute potency of the compound into account. Thus, we initially observed many connections with significant p-values, but when looking at the sensitivity distributions, the compounds did not exhibit patterns that we were interested in (e.g., all unresponsive).

This issue was addressed by several steps: 1) filtering out compounds which did not show a desirable range of sensitivities. Three filtering criteria were used here: (i) for sensitive or resistant enrichments, the AUC ≤ 3.5 or AUC ≥ 5.5 (AUC=7 corresponds to no compound effect) for at least one CCL respectively, (ii) $AUC_{highest} - AUC_{lowest} \geq 3.0$, (iii) at least one CCL harboring the genetic feature was tested with the compound; 2) a parametric chi-squared test of homogeneity to account for the absolute potency of each compound in relation to the distribution of genetic alterations was performed, producing a list of potential cancer dependencies with statistical significance with the desired compound sensitivity performance.

The maximum (worst case) of the two p-values from the non-parametric and parametric tests was squared to obtain a significance score for subsequent multi-test correction and ranking. To control for multiple hypothesis testing, the Benjamini-Hochberg procedure[30] was applied within each family of hypotheses (different genetic or lineage features sharing a measured AUC distribution). This generated q-values (adjusted squared maximum p-values). A false-discovery rate (FDR) cutoff of $q < 0.25$ was applied for Suppl. Table S5.2 (digital), and a more stringent one for the CTRP.

CCL sub-population analysis: As not all lineage-based sub-populations and excluded sub-populations actually required consideration, we checked each lineage and sub-lineage by itself, as well as combinations of different lineages and exclusions before including them in our analysis. Two steps were taken to assess whether certain sub-populations of CCLs have non-compound-specific sensitivity characteristics that confound our enrichment analyses: 1) identify these sub-populations, regardless of specific genetic lesions; 2) determine the influence on the enrichment analyses if various sub-populations are analyzed separately vs together.

Identification of potentially confounding sub-populations: Komolgorov-Smirnov (K-S) tests were performed for qualitative (categorical exclusions), asking whether the exclusion of growth conditions (adherent, suspension, mixed) or individual lineages had significant effects on AUC distributions. Excluding suspension CCLs and excluding hematopoetic CCLs was found to be significant ($p_{KS} < 3.67\times10^{-9}$ and $p_{KS} < 2.34\times10^{-10}$, respectively).

In order to determine a threshold to identify CCLs with statistically large numbers of mutations, it was assumed that these CCLs would have a different sensitivity

distribution than all other CCLs. The difference between distributions of CCL sensitivity between the *i* most-frequently-mutated CCLs and the remainder was determined using a collection of *i* two-distribution K-S tests, subject to Bonferroni correction, was used. 33 CCLs were determined to have 'genes mut high', meaning they harbored a fraction of mutations greater than the most significant K-S test ($p_{KS} < 1.63\times10^{-11}$).

To identify CCLs that were sensitive to many compounds, a similar analysis was done, but many K-S test significance values exceeded machine precision. 32 CCLs were thus termed 'frequently sensitive', meaning that they were sensitive to more than 25% of the small molecules. These subsets of CCLs were excluded for parts of the enrichment analyses.

To include a lineage or sub-lineage in the analysis, more than 3 CCLs present in the dataset had to be of that lineage or sub-lineage. For lineage and exclusion combination analysis, combinations were included if at least 3 CCLs passed one of the four exclusion criteria and were present in the respective lineage or sub-lineage. Of the lung lineage, the following sub-lineages were tested: adenocarcinoma, broncheoalveolar carcinoma, large cell carcinoma, non-small cell carcinoma, small-cell carcinoma, and squamous cell carcinoma; and of the hematopoietic lineage: acute myeloid leukemia, diffuse large B-cell lymphoma, plasma cell myeloma.

Global Analysis: Including all lineage/sub-lineage connections, and exclusions across all datasets, 397,270 enrichments were identified. Each pair of scores was considered using each exclusion and for each compound-gene connection. The score-pairs were conceptualized as falling into interpretable regions of a scatterplot indicating if connections were improved, preserved, or diminished upon running the corresponding

exclusion experiment. The number of score-pairs for each region in the scatterplot was compared to a randomly permuted score matrix. Only excluding suspension or hematopoietic CCLs is independent of the sensitivity data, and thus, only these two exclusions were used to qualify connections for the global analysis. If a connection improved or preserved when both suspension and hematopoietic lines were excluded, the best-scoring connection was used. If either exclusion reduced the connection, the reduced score was used, and if either exclusion caused the connection to become insignificant, it was removed. Additionally, connections with contradictory scores between exclusions and/or the primary analysis were removed. If the same compound-gene connection was enriched in several datasets, only the best-scoring connection was kept. This resulted in a set of connections dominated (in number) by connections with relatively weaker q-value scores. Therefore, connections with <3 mutations, <3 sensitive mutants, or <3 unresponsive mutants, as well as connections where more than half of CCLs tested were mutant CCLs, were excluded. The global analysis was thus run on 108,635 candidate compound-gene connections.

To retain the strongest gene-compound connections and protect against type I errors, the optimal threshold for a more stringent q-value was determined. This was done by varying the q-value between 0 and 0.25 and observed the changes in the number of connections in the fraction of remaining connections. These changes were small (i.e., relative stability in the connection list was achieved) at a cutoff of $q<0.025$, which was used to determine 16,667 distinct compound-gene connections that qualified for further analyses.

The frequency, sum of scores, and average scores for each gene and compound was

calculated individually in both sensitive and unresponsive directions (Suppl. Table S5.3a-d (digital)). The number of overlapping genes and compounds and their significance (by hypergeometric distribution) was computed for each pair of compounds and genes, respectively (Suppl. Table S5.3e-h (digital)). Subsequently, complete-linkage clustering analysis on the compounds was performed using a cosine similarity distance based on the presence or absence of a connection between each compound and gene (binary calls) (Suppl. Table S5.3i (digital), Suppl. Figure S5.2). For each gene where the compound connection score with the gene ≠ 0, a weight was calculated within each cluster. This weight was proportional to the fraction of compounds to which the gene connection, e.g., if a gene connected to all (half of all) compounds in a cluster, then the weight for that gene was 1 (0.5). The mean q-score of the gene across all compounds in the cluster was multiplied with the weight per gene, and the products summed. A random score for a cluster of size n over 100 iterations was computed to determine the significance of the calculated score. The reported score (Suppl. Table S5.3i (digital)) is $(s_i - \mu_{random,n})/\sigma_{random}$. All non-zero genes that were associated with each reported cluster of compounds and their respective weights (Suppl. Table S5.3j (digital)) are also reported. The same was done for the lineage-specific analysis where 46,175 total qualified connections (involving >2 mutants, >2 examples (across all datasets) were used, resulting in 12,518 distinct gene-compound connections.

References

(1) Reprinted from *Cell*, *154* (5), Basu, A. *et al.*, "An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules", 1151-1161. Copyright (**2013**), with permission from Elsevier.

(2) Brown, E. J. *et al.* "Control of p70 s6 kinase by kinase activity of FRAP in vivo." *Nature* **1995**, *377*, 441-446.

(3) Liu, J. *et al.* "Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP- FK506 complexes." *Cell* **1991**, *66*, 807-815.

(4) Pan, H., Cao, J. & Xu, W. "Selective histone deacetylase inhibitors." *Anti-Cancer Agents Med. Chem.* **2012**, *12*, 247-270.

(5) Cheung, H. W. *et al.* "Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer." *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 12372-12377.

(6) Smalley, K. S. "PLX-4032, a small-molecule B-Raf inhibitor for the potential treatment of malignant melanoma." *Curr. Opin. Invest. Drugs (BioMed Cent.)* **2010**, *11*, 699-706.

(7) Prahallad, A. *et al.* "Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR." *Nature* **2012**, *483*, 100-103.

(8) Straussman, R. *et al.* "Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion." *Nature* **2012**, *487*, 500-504.

(9) Corcoran, R. B., Settleman, J. & Engelman, J. A. "Potential therapeutic strategies to overcome acquired resistance to BRAF or MEK inhibitors in BRAF mutant cancers." *Oncotarget* **2011**, *2*, 336-346.

(10) Yoon, J., Koo, K. H. & Choi, K. Y. "MEK1/2 inhibitors AS703026 and AZD6244 may be potential therapies for KRAS mutated colorectal cancer that is resistant to EGFR monoclonal antibody therapy." *Cancer Res.* **2011**, *71*, 445-453.

(11) National Cancer Institute. *NLM Identifier: NCT01229150. Randomized Phase II Study of AZD6244 MEK-Inhibitor With Erlotinib in KRAS Wild Type and KRAS Mutant Advanced Non-Small Cell Lung Cancer.* , <http://clinicaltrials.gov/ct2/show/record/NCT01229150 > (accessed: 06/18/2012).

(12) Jänne, P. A. *et al.* "Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study." *Lancet Oncol.* **2013**, *14*, 38-47.

(13) Arteaga, C. L. "EGF receptor mutations in lung cancer: from humans to mice and maybe back to humans." *Cancer Cell* **2006**, *9*, 421-423.

(14) Vervoorts, J., Lüscher-Firzlaff, J. & Lüscher, B. "The ins and outs of MYC regulation by posttranslational mechanisms." *J. Biol. Chem.* **2006**, *281*, 34725-34729.

(15) Ran, Z. H., Zou, J. & Xiao, S. D. "Experimental study on anti-neoplastic activity of epigallocatechin-3-gallate to digestive tract carcinomas." *Chin. Med. J. (Beijing, China, Engl. Ed.)* **2005**, *118*, 1330-1337.

(16) Ju, J. *et al.* "Inhibition of intestinal tumorigenesis in Apcmin/+ mice by (-)-epigallocatechin-3-gallate, the major catechin in green tea." *Cancer Res.* **2005**, *65*, 10623-10631.

(17) Acosta, J. C. *et al.* "Chemokine signaling via the CXCR2 receptor reinforces senescence." *Cell* **2008**, *133*, 1006-1018.

(18) Chen, Z. *et al.* "A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response." *Nature* **2012**, *483*, 613-617.

(19) Sartore-Bianchi, A. *et al.* "Multi-determinants analysis of molecular alterations for predicting clinical benefit to EGFR-targeted monoclonal antibodies in colorectal cancer." *PLoS One* **2009**, *4*, e7287.

(20) Weiwer, M. *et al.* "Development of small-molecule probes that selectively kill cells induced to express mutant RAS." *Bioorg. Med. Chem. Lett.* **2012**, *22*, 1822-1826.

(21) Skehan, P. *et al.* "New colorimetric cytotoxicity assay for anticancer-drug screening." *J. Natl. Cancer Inst.* **1990**, *82*, 1107-1112.

(22) Dixon, S. J. *et al.* "Ferroptosis: an iron-dependent form of nonapoptotic cell death." *Cell* **2012**, *149*, 1060-1072.

(23) Corcoran, R. B. *et al.* "Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes tumor regressions in KRAS mutant cancer models." *Cancer Cell* **2013**, *23*, 121-128.

(24) Garnett, M. J. *et al.* "Systematic identification of genomic markers of drug sensitivity in cancer cells." *Nature* **2012**, *483*, 570-575.

(25) Hanahan, D. & Weinberg, R. A. "Hallmarks of cancer: the next generation." *Cell* **2011**, *144*, 646-674.

(26) Luo, J. *et al.* "A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene." *Cell* **2009**, *137*, 835-848.

(27) Dancik, V. *et al.* "Connecting Small Molecules with Similar Assay Performance Profiles Leads to New Biological Hypotheses." *J. Biomol. Screening* **2014**.

(28) Barretina, J. *et al.* "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity." *Nature* **2012**, *483*, 603-607.

(29) Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. *Introduction to algorithms*. 2nd edn, (MIT Press :McGraw-Hill, **2001**).

(30) Benjamini, Y. & Hochberg, Y. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *J. R. Statist. Soc. B* **1995**, *57*, 289-300.

Page intentionally left blank.

Chapter Six

Mining the cancer dependency resource for novel dependencies in recurrently-mutated and fusion genes

The Chemical Biology Program and Platform led by Dr. Josh Bittker, Dr. Jaime Cheah, Matthew Coletti, Victor Jones, Edmund Price, and Christian Soule collected the data for the Phase 2 profiling effort. Dr. Paul Clemons and Nicole Bodycombe processed the data. Dr. Michelle Stewart and I collaborated on this project. Together, we prepared the input list for recurrent mutations. Dr. Stewart lifted the TCGA files that were in hg18 to hg19 and I annotated all mutations in our TCGA and CCLE data files with advice from Dr. Lee Lichtenstein. We both analyzed the overlap and determined the list of recurrent features to test. Dr. Stewart ran the enrichment analysis of recurrent genes. Dr. Stewart and I together picked and retested the CCLs for their response to PI3K and AKT inhibitors. I compiled the list of fusion genes and Dr. Clemons ran the enrichment analysis for them.

Expanding the CTRP dataset

The 242 cancer cell lines (CCLs) profiled in the last chapter already gave rise to a number of hypotheses for novel cancer dependencies. Increasing numbers of CCLs and compounds profiled will allow more insight into possible dependencies. The Chemical Biology Program and Platform at the Broad Institute have engaged in a second profiling effort to test 893 human CCLs in 1,536-well format with an Informer Set of 495 single agent small molecules and 50 combinations (Suppl. Table S6.1 (digital)), the data of which will be publicly available on the CTRP (http://www.broadinstitute.org/ctrp). Mining this dataset has uncovered novel challenges in interpreting the data and overcoming these has led to new analyses and promising hypotheses.

Frequencies of small-number-mutant enrichments

When interpreting enrichments from the CCL profiling, the number of mutated CCLs per enrichment is an important feature to consider for analysis. Many of the enrichments are based on a small number of cell lines that have the enriched feature. For example, of the 397,270 enrichments in Suppl. Table S5.2 (digital), 97,781 have a single mutant CCL and 87,595 have two mutant CCLs (Figure 6.1a). Often, the CCLs with these mutations are not the only sensitive ones and little is known about the impact the specific mutations have. Additionally, for many of the enriched features, the mutations occur in the same gene, but not at the same amino acid position. Therefore, it is not guaranteed that they have the same biological impact. These factors make it hard to interpret whether mutations in the enriched gene are the deciding factor for sensitivity.

Enrichments with less than three mutant CCLs were therefore excluded for the online CTRP. Still, 50% of the remaining enrichments are with smaller numbers of mutant-CCLs (3-5, Figure 6.1b), and the interpretation challenge remains, especially with mutations that lack understanding of biological function. While these might still harbor interesting insights which could be explored by trial and error, systematic prioritization of enrichments could expedite the discovery of relevant hypotheses.

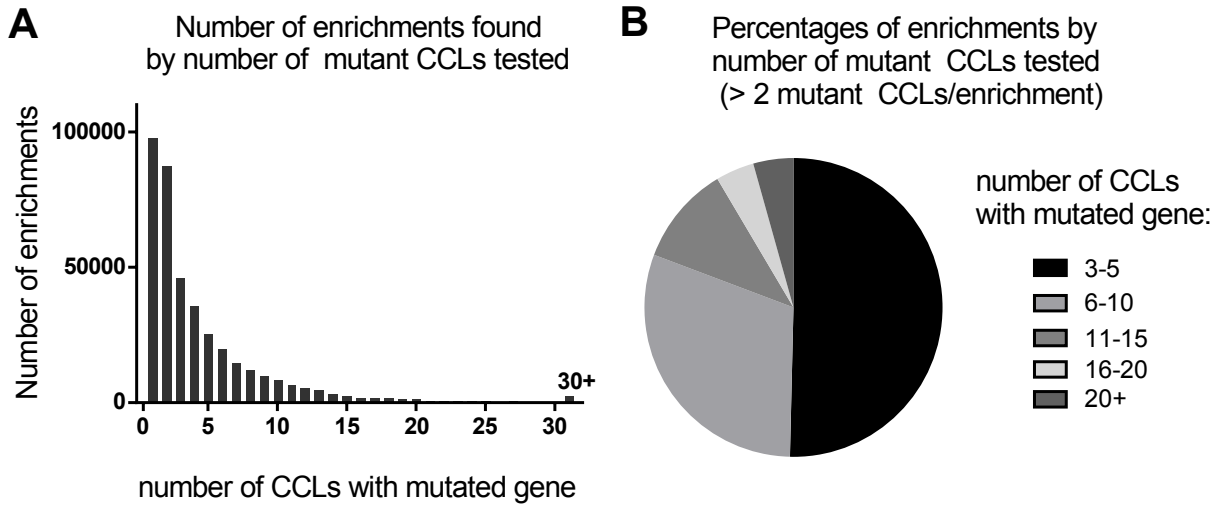


Figure 6.1. a) number of enrichments generated in the CTRP shown by number of the respective mutant CCLs tested. b) percentages of enrichments generated in the CTRP shown number of respective mutant CCLs tested (out of all enrichments for features where >2 CCLs with the feature were tested).

Finding biologically relevant mutations

With this challenge in mind, we thought about ways to prioritize enrichments in the data. The CCLE analyzed >1,600 genes known to be frequently mutated in cancer.[1] While some of the mutations in these genes are well-known oncogenes (e.g. *KRAS* G12 mutations), the function of many of these mutations is unknown. Some of these mutations might have a direct impact on cancer biology (driver mutations), but many might also be passenger mutations that do not contribute to tumor development.[2,3] The

CCLE also captured germline variants, which might account for some of these single mutations. Additionally, given that many CCLs have been cultured for decades, it is also likely that they acquired mutations over time that patient tumors have not had the time to acquire. Consequently, it is known that CCLs have different genetic characteristics than patient tumors.[4-7] Therefore, one challenge when working with the mutations in CCLs is defining which mutations are relevant to tumor biology. In order to sort out which mutations are biologically relevant, and which ones are not, we decided to restrict the mutations we would look at in two ways: 1) mutations had to have been found in patients and 2) mutations had to be recurrent.

CCLE and TCGA dataset overlap

The rationale behind limiting ourselves to mutations that are present in patients is that if a mutation was caused by extended cultivation *in vitro*, but has no impact on actual tumor biology, it is very unlikely that this mutation would occur in patients as well. The patient-derived data was taken from the The Cancer Genome Atlas (TCGA). These samples are taken directly from patients and matched with non-cancerous tissues of each patient to determine somatic mutations.[8] Therefore germline variants would not be captured. The mutations in these samples should be more relevant to cancer biology, even though some confounding factors, such as random passenger mutations, will also exist in them.

We first set out to determine the overlap of mutations that occur between the CCLE and the TCGA datasets. For this purpose, we downloaded all available TCGA datasets using the Broad Institute's Cancer Genome Analysis (CGA) group Firehose

tool (http://www.broadinstitute.org/cancer/cga/firehose) as well as the mutational data of the CCLE oncomap and hybrid capture datasets.[1,9] In order to make sure that all mutations were annotated using the same human genome (hg) reference batch and protein call, we removed all annotated protein mutation information. We then transferred hg18 data to hg19 data using the University of Santa Cruz's LiftOver Tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver).[10-12] Subsequently, we annotated all mutations into protein changes using the chromosome, start and end positions, mutated allele and reference allele for all samples. This was done using the Broad Institute's Oncotator Tool (www.broadinstitute.org/cancer/cga/oncotator).[13] We also filtered the TCGA data to only include the 1,667 genes that are contained in the CCLE.

We then determined how many mutations overlapped using two ways of classifying mutations: by exact protein change (e.g. *KRAS* G12D) and by amino acid location (e.g. *KRAS* G12, mutation could result in any amino acid) (Table 6.1). The percentage of CCLE mutations that overlapped with the TCGA is low, but does increase when recurrent mutations are examined. The CCLE dataset has many more mutations per sample, which could be due to the capturing of private germline variants, and extended growth *in vitro*. Therefore, limiting the mutations to those present in patient data could exclude some of these confounding factors.

Recurrent mutations

The second restriction we used was that mutations had to be recurrent. It has been shown that some CCLs are hypermutated and it has been suggested that care should be exercised in picking CCLs as models for tumor studies.[4]

Table 6.1. overlap between the CCLE and TCGA datasets by protein change and location

number of mutations	TCGA (4352 samples)	CCLE (957 samples)	overlap if > x in both	% CCLE data	overlap if in TCGA	% CCLE data
protein change						
all	126835	53564	3940	7.4		
> 1	6902	4111	346	8.4	758	18.4
> 2	1339	1093	149	13.6	328	30.0
location change						
all	117466	51516	8254	16.0		
> 1	12342	4973	515	10.4	1327	26.7
> 2	2370	1237	188	15.2	463	37.4

Many studies have been done on the functional impact of recurrent mutations, and recurrent mutations are considered to be biologically relevant.[14,15] These mutations are not only random mutations, and, if they also happen in patient samples, they were most likely were selected for by a factor of tumor biology.

When studying recurring mutations in the CCLE sample, it became clear that there are a number of recurrent mutations that occur in a majority of the CCLs (Figure 6.2a). However, this frequency of recurrent mutations is not reflected in the TCGA data (Figure 6.2b) and many of these mutations do not occur in that dataset at all. Some of these could be due to sequencing artifacts, which is especially supported given many of them are splicing mutations, insertions, or deletions, which are known to be challenging to capture correctly.[16] Given that many CCLs have been cultured for decades, it is also possible that they acquired additional mutations over time that patient tumors have not had the time to acquire, or that part of the selection that lets these cells grow *in vitro* requires them to have more mutations in certain genes. Many of the recurrent TCGA

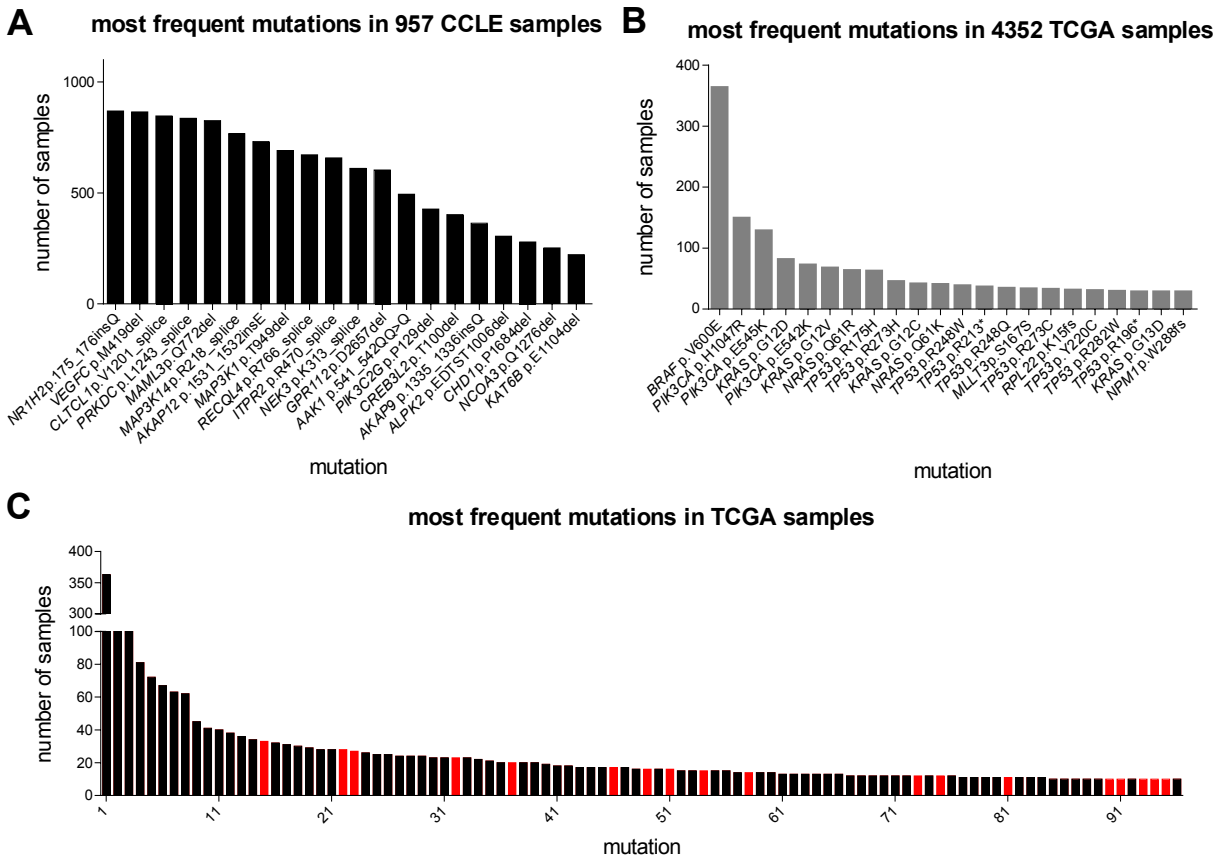


Figure 6.2. a) most frequent mutations in CCLE samples. b) most frequent mutations in TCGA samples. c) all frequently mutated (> 10) samples in TCGA. Red bars correspond to mutations not present in the CCLs.

mutations are well-known to drive cancer. With few exceptions (*e.g.*, the silent *MLLT3* S167S mutation, or *NPM1* W288fs mutations), these are also present in CCLs (Figure 6.2c). Therefore, the CCLE dataset covers most of the main recurrent mutations found in tumors and we also chose to restrict our dataset to recurrent mutations.

Enrichment analysis with recurrent mutations and fusion genes

Since the CCLs seem to have many mutations that are not present in patient samples, prioritizing mutations which overlap between the datasets might give us more relevant hypotheses. Additionally, overlapping, recurrent mutations seem to be

biologically very relevant. We therefore decided to run the enrichment analysis with two new sets of features: 1) all genes with those mutations that also exist in patient data and 2) all genes with specific recurrent (present in > 2 CCLs tested) mutations in CCLs that are also recurrent (> 2 samples) in patient data. We again considered both protein change and location as two different sets of features. As CCLs from the hematopoietic/lymphoid lineage were generally very sensitive to compounds (Chapter 5), we decided to exclude them for the purpose of this enrichment.

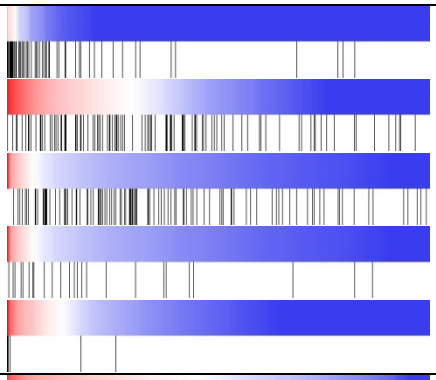
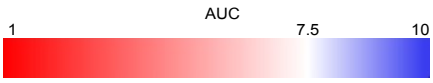
One caveat in restricting the dataset in this manner is that there might be more mutations in patients that are actually represented in the CCLE dataset, but that are being excluded with the current restriction. As more patient samples become available, it might therefore be useful to repeat this analysis, as connections with these other biologically relevant mutations could exist.

Another dataset that was not considered in the original set was that of fusion genes. Fusion genes are often involved in cancers and can be strong drivers.[17] Unfortunately, full characterization of fusion genes was not available for the CCLs. The Sanger Catalogue of Somatic Mutations in Cancer (COSMIC, www.cancer.sanger.ac.uk) has characterized CCLs for some fusion genes that are known to drive cancer.[18,19] I decided to incorporate this data into the enrichment analysis for those fusion genes for which data existed.

Known dependencies are found in the enrichment

The enrichment analysis for these three new sets of features gave a large number of possible new cancer dependencies (Table 6.2, Suppl. Table S6.2 (digital)).

Table 6.2. selected enrichments from new analysis

compound	compound target	enrichment	FDR q-value	heatmap
vemurafenib	BRAF V600E	*BRAF* V600E	$1.18 \cdot 10^{-21}$	
selumetinib	MEK 1/2	*KRAS* G12	$2.03 \cdot 10^{-5}$	
PD318088	MEK1/2	*KRAS* G12	$9.91 \cdot 10^{-5}$	
PD318088	MEK1/2	*NRAS* Q61	$1.14 \cdot 10^{-4}$	
gefitinib	EGFR	*EGFR* location 746	$7.6 \cdot 10^{-3}$	
GDC-0941	PI3K	*PIK3CA* E542K	$1.25 \cdot 10^{-3}$	
BYL-719	PI3K-α	*PIK3CA* E542K	$1.04 \cdot 10^{-3}$	
AZD6482	PI3K-β	*PIK3CA* H1047R	$3.10 \cdot 10^{-2}$	
MK-2206	AKT	*PIK3CA* E542K	$2.13 \cdot 10^{-2}$	
clofarabine	POLA1/2/E; RRM1	*CDC25C* K244fs	$8.55 \cdot 10^{-3}$	
NVP-231	CERK	*CDC25C* K244fs	$7.86 \cdot 10^{-3}$	
austocystin D	cytotoxic/ P450 activation	*ACVR2A* K435fs	$2.34 \cdot 10^{-2}$	
			legend:	

The analysis did discover known dependencies of recurrent mutations. For example, *BRAF* V600E-mutant CCLs are very sensitive to vemurafenib, an FDA-approved V600E-targeted BRAF inhibitor.[20] Other BRAF inhibitors (e.g. PLX-4720, GDC-0879, dabrafenib) also show this enrichment of V600E-mutant CCLs (Suppl. Table S6.2 (digital)). CCLs with mutations in *KRAS* amino acid position 12 are enriched for the MEK 1/2 inhibitor selumetinib, a compound that has shown activity in patients with *KRAS*-mutant cancer.[21] Both *KRAS* G12 and *NRAS* Q61-mutated CCLs are also

enriched for another MEK1/2 inhibitor, PD318088.[22] MEK inhibitors have shown activity in patients with *KRAS*- or *NRAS*-mutant cancers.[23,24] Additionally, CCLs with mutations (deletions) in amino acid position 746 of *EGFR*, show enhanced sensitivity to EGFR inhibitors such as gefitinib, as reported.[25,26]

Some compounds are only effective in CCLs with specific mutations

When looking at the data, it becomes apparent that some compounds enrich more in CCLs with specific mutations rather than all mutants. For example, the PI3K-pan and α-inhibitors GDC-0941 and BYL-719, show enrichment with *PIK3CA* E542K-mutated CCLs, but not with other PI3K mutations (Table 6.2, Figure 6.3a). In contrast, the PI3K β-inhibitor AZD6482 shows enrichment for *PIK3CA* H1047R-mutated cell lines (Table 6.2, Figure 6.3a). GDC-0941 and BYL-719 are in clinical trials for a variety of cancers[27] and AZD6482 has been tested for safety and tolerability.[28] While patients in other clinical trials have been stratified by *PIK3CA* mutations in exon 9 (e.g. E542K, E545K), or exon 20 (e.g. H1047R),[29] the specific locations have not been distinguished and neither compound has been connected to specific *PIK3CA* mutations. If this hypothesis holds up, these compounds should preferentially be given to patients with the respective *PIK3CA* mutations. Another compound that shows enrichment with *PIK3CA* E542K mutant CCLs is the AKT inhibitor MK-2206 (Table 6.2, Figure 6.3).[30] PI3K and AKT are in the same pathway.[31,32] It is therefore possible that certain *PIK3CA* mutations influence AKT signaling in different ways.

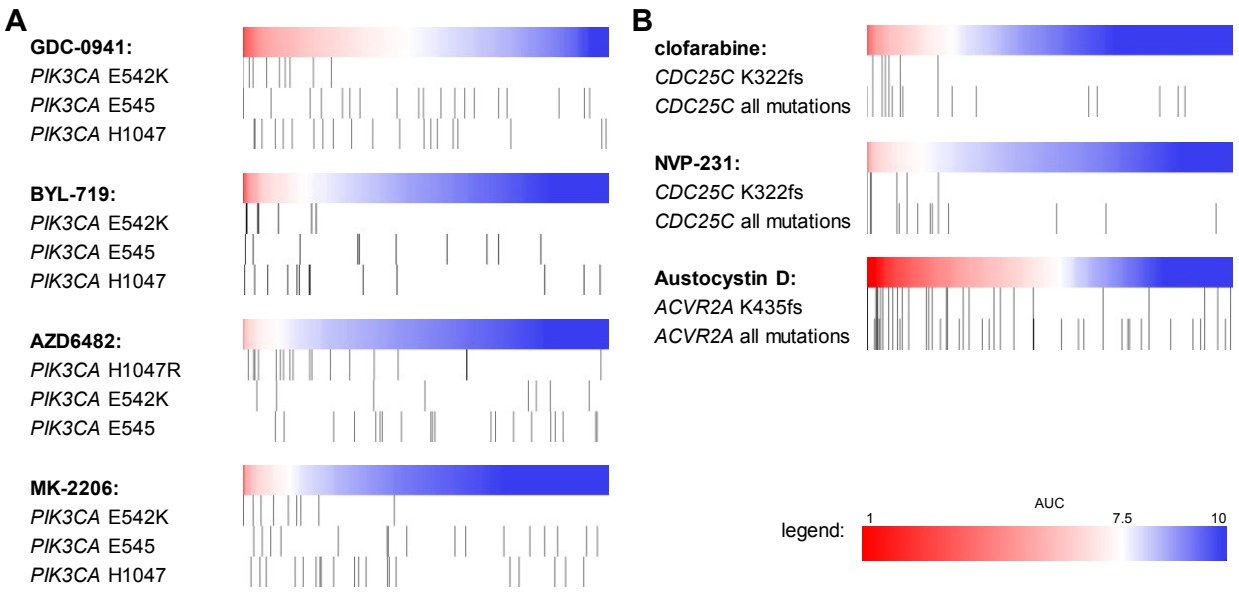


Figure 6.3. Heatmaps for recurrent and all mutations of connections to *PIK3CA* (a) and *CDC25C*, *TNKS2*, *ACVR2A* (b).

More hypotheses that might be of high interest to follow-up on can be found in the new enrichment analysis. For example, CCLs harboring 244fs mutations in *CDC25C* are very sensitive to clofarabine, a DNA-polymerase and ribonucleotide reductase inhibitor (Table 6.2, Figure 6.3b).[33] This compound is already FDA-approved for a variety of leukemias.[34] Like DNA polymerase, CDC25C is involved in cell division[35] and this connection could point towards a dependency of *CDC25C* 244fs-mutant CCLs on mitotic processes. Similarly, CCLs harboring this mutation are also very sensitive to NVP-231, a ceramide kinase inhibitor.[36] Ceramide kinase produces ceramide-1-phosphates which have been shown to stimulate DNA synthesis.[37,38] Another connection that the new enrichment analysis suggests is that of *ACVR2A* 435fs-mutant CCLs to austocystin D (Table 6.2, Figure 6.3b). This compound has been shown to be selectively activated into a cytotoxic derivative in CCLs with high cytochrome P450 expression.[39] Activin signaling has been shown to interact with different cytochrome

P450 enzymes, both in activating and inactivating functions.[40,41] It is possible that the frameshift mutations in *ACVR2A* influence the expression of the enzymes responsible for austocystin D activation and thus cause selective killing of these cells.

Retesting *PIK3CA*-mutant sensitivity

In order to test the *PIK3CA*-mutant sensitivity hypotheses, we retested a panel of *PIK3CA* mutant CCLs in 384-well format against these inhibitors. We added several other PI3K/AKT inhibitors to the test: BKM120, a pan PI3K inhibitor, CH5132799, a PI3K-α inhibitor, and GDC-0068, an AKT inhibitor.[42-44] We also added one previously untested *PIK3CA* E542K-mutant CCL, BT483. We tested each cell line at 2000 cells/well, a density that was similar to that used in 1,536 well-format, and at an optimized density determined during assay development (Table 6.3).

In retesting, the trend of *PIK3CA* E542K mutants being more sensitive than other *PIK3CA* mutants to pan-PI3K or PI3K-α inhibitors was confirmed (Figure 6.4a-d, Suppl. Figure S6.1). However, the data does not support the original result for AKT inhibitors strongly (Suppl. Figure S6.2). The trend of *PIK3CA* H1047R-mutant CCLs to be more sensitive to the PI3K-β inhibitor AZD6482 was also confirmed (Figure 6.4e, f). However, at optimized densities, there are several non-*PIK3CA* H1047R-mutant CCLs that are more sensitive than the mutant CCLs. This could mean that the hypothesis is incorrect and other factors, such as seeding density, influence sensitivity which needs to be investigated in further follow-up studies.

Table 6.3. CCLs used for PI3K-retest and densities used.

CCL	Lineage	*PIK3CA* mutation	optimized density (cells/well)
BC3C	urinary tract	E545K	500
BT20	breast	H1047R	1000
BT483	breast	E542K	2000
CAL29	urinary tract	H1047R	500
CAL51	breast	E542K	1000
HCC202	breast	E545K	2000
HCT116	large intestine	H1047R	500
HCT15	large intestine	E545K	500
HCT8	large intestine	E545K	250
HGC27	stomach	E542K	500
HT1197	urinary tract	E545K	500
IM95	stomach	E542K	1000
LS180	large intestine	H1047R	1000
MDAMB361	breast	E545K	2000
MDAMB453	breast	H1047R	1000
MKN1	stomach	E545K	500
NCIH1341	lung	E542K	2000
NCIH460	lung	E545K	250
NCIH508	large intestine	E545K	1000
RKO	large intestine	H1047R	1000
SNU407	large intestine	H1047R	1000
SNUC5	large intestine	H1047R	2000
SW948	large intestine	E542K	1000
T84	large intestine	E542K	2000
VMCUB1	urinary tract	E542K	500

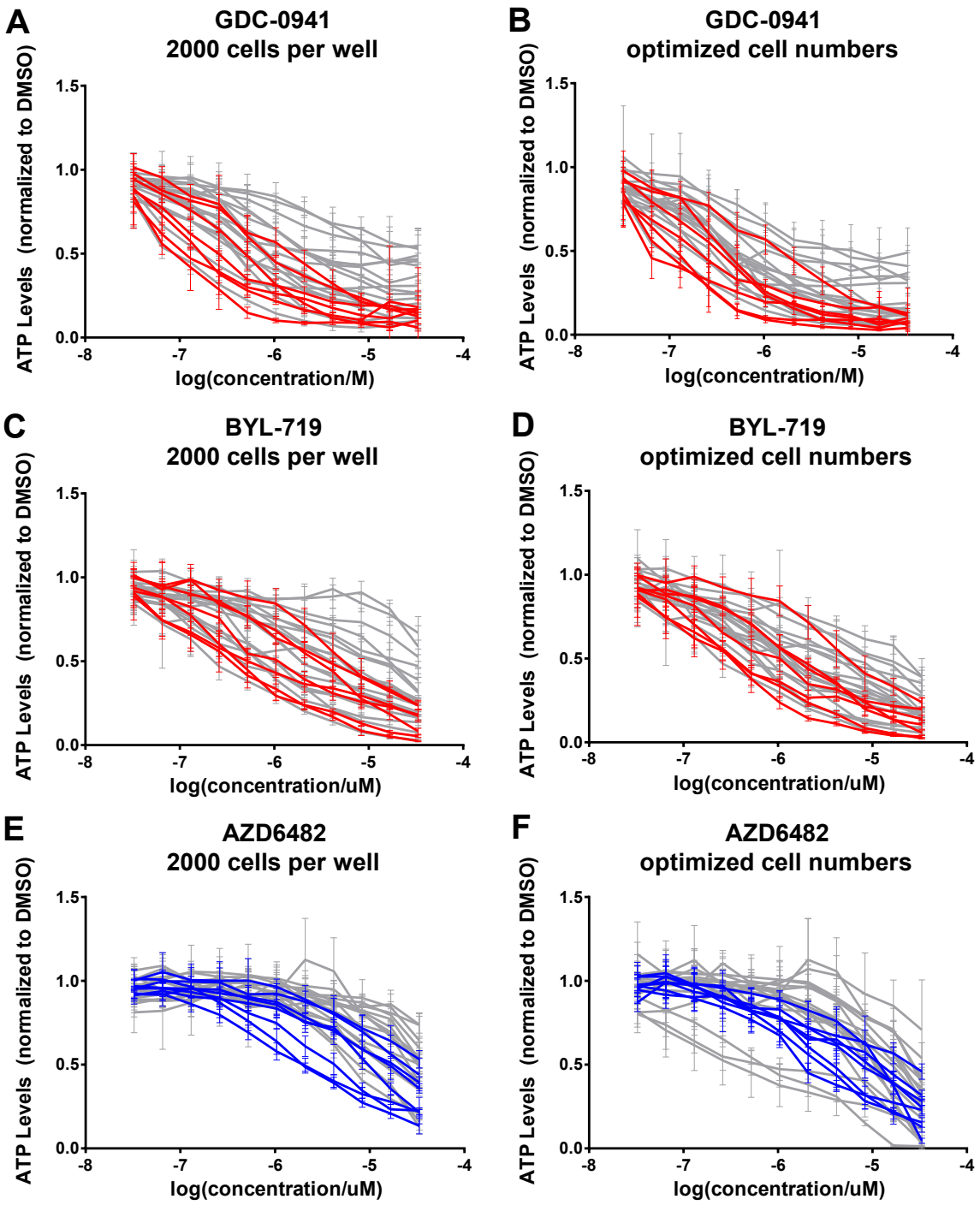


Figure 6.4. Dose-response curves of PI3K inhibitor re-test in 384-well format. All values are shown and were generated in two independent experiments run in triplicate (values are mean ± SD). Red dose-response curves are of *PIK3CA* E542K-mutant CCLs, blue dose-response curves are of *PIK3CA* H1047R-mutant CCLs. a-b) dose-response curves for GDC-0941 in mutant *PIK3CA* CCLs at 2000 cells/well (a) or optimized densities (b). c-d) dose-response curves for BYL-719 in mutant *PIK3CA* CCLs at 2000 cells/well (c) or optimized densities (d). e-f) dose-response curves for AZD6482 in mutant *PIK3CA* CCLs at 2000 cells/well (e) or optimized densities (f).

The hypotheses look promising and could be used to increase the benefit of PI3K inhibitors in the clinic by treating patients with the correspondingly more sensitive mutations. It is important to note that the dose-response curves for many compounds were less separated between sensitive and nonresponsive CCLs at optimized seeding densities. It has been shown that mutant *PIK3CA* can allow CCLs to overcome normal contact-inhibitory signals.[45] It is possible that overgrowth of CCLs changes their sensitivity to the compounds. Therefore, more studies are necessary to confirm the hypothesis and bring possible benefit to patients. Genomic cell line engineering, using tools such as CRISPR,[46,47] could help prove a direct cause of the mutation. Additionally, since these compounds are in clinical trials, retroactive analyses of trial results might be refined to specific mutations, rather than mutational status of the whole gene.

Fusion gene enrichment

The enrichment analysis of the fusion genes also added new hypotheses. *NPM1/ALK*-translocated CCLs were very sensitive to treatment with NVP-TAE684 and crizotinib, two ALK inhibitors (Table 6.4, Suppl. Table S6.3 (digital)).[48,49] Crizotinib has been shown to be effective in clinical trials in patients with ALK translocations.[50] Therefore, fusion genes are a valid feature to examine in CCL profiling. Because the annotation of fusion genes within CCLs is still sparse, not many other hypotheses arise from this analysis. *NPM1/ALK*-translocated CCLs are also sensitive to the CDK inhibitor PHA-793887.[51] This is contrary to a previous study showing that *NPM1/ALK*-positive CCLs from anaplastic large cell lymphoma are less sensitive than *ALK*-negative CCLs to another CDK inhibitor flavopiridol.[52] This discrepancy could arise because the authors

Table 6.4. selected enrichments from fusion gene analysis

compound	compound target	enrichment	p-value	heatmap
NVP-TAE684	ALK	NPM1/ALK	$1.15 \cdot 10^{-8}$	
crizotinib	ALK	NPM1/ALK	$2.78 \cdot 10^{-8}$	
PHA-793887	CDK	NPM1/ALK	$3.33 \cdot 10^{-7}$	
BRD-K92856060	unknown	EWSR1/FLI	$8.65 \cdot 10^{-8}$	
			legend:	AUC 1 7.5 10

did not study many other CCLs, or because PHA-793887 has a different activity profile, causing it to have different effects. However, PHA-793887 has been shown to have significant hepatotoxicity in clinical trials,[53] so another compound with a similar inhibitory profile, but no hepatotoxicity would have to be found in order to benefit from this hypothesis. Additionally, *EWSR1/FLI1*-translocated CCLs are very sensitive to a compound called BRD-K92856060. This compound does not have an annotated mechanism of action, but a PubChem search (CID: 2945648) shows that it had bioactivity in several assays, most notably in 14 assays related to microphthalmia-associated transcription factor (MITF).

When studying the annotations that are in COSMIC, it is obvious that many CCLs have not yet been annotated properly for fusion genes. For example, the well-known fusion gene *BCR/ABL*, which causes chronic myelogenous leukemia (CML)[54-56] is not annotated for any CCL, even though some CCLs are of the CML sublineage, according to CCLE. Therefore, this analysis should be repeated with full annotations of fusion genes once available, as more relevant hypotheses might arise from the process.

Conclusion

In conclusion, CCL profiling can generate many hypotheses regarding cancer dependencies. Some of these hypotheses, if tested accordingly, could change patients' lives quickly, as the compounds involved are already in the clinic. Many more hypotheses can be uncovered when carefully searching the data. However, care should be exercised when mining the data for relevant hypotheses, as CCLs have many confounding mutations. As all hypotheses will need to be confirmed in costly experiments, picking the most promising ones is highly desirable. Selecting enrichments with strong p- and q-values and prioritizing those with known relevant mutations could lead to faster success. Additionally, apart from retesting the effects in more CCLs, directed follow-up experiments (*e.g.*, using genomic editing) can help establish direct correlation between mutations and their effects in a quick and straightforward manner. Moving forward, we plan to follow-up on selected hypotheses using these assay systems.

Experimental Section

Materials and Methods

Cell Lines and Reagents: cells were purchased from the Broad Institute Biological Samples Platform or ATCC (Manassas, VA, USA). All CCLs were grown at 37°C/5% CO_2 in their specified medium (Suppl. Table S6.1 (digital)). All CCLs were tested for mycoplasma using a Takara PCR Mycoplasma Detection Kit (Otsu, Japan). Media components and supplements were purchased from ATCC (VA, USA), Cellgro

(VA, USA), Life Technologies (Carlsbad, CA, USA), EMD Millipore Chemicals (Billerica, MA, USA), or Invitrogen (Carlsbad, CA, USA).

Assay Treatments and Readouts: as described in Chapter 5. In addition: Plate Reader settings used were: luminescence settings 1,536-well: Perkin Elmer (Waltham, MA, USA) Viewlux 1430 uHTS Microplate Imager. Compounds were acoustically-transferred using a Labcyte Echo 555 (1,536-well) or pin-transferred using a CyBio (Jena, Germany) Cy-Bi Well Vario (384-well).

Assay Protocols

Cancer Cell Line (CCL) Profiling: Cells were plated at 500 cells/well, 6 μL/well in 1,536-well opaque, moat-well, white assay plates and incubated overnight at 37°C/5% CO_2. Compound stocks were plated in 1,536-well format in 16-pt, 2-fold concentration ranges defined by literature review. Compounds (20 nL) were acoustically-transferred into duplicate assay plates and incubated for 72h. The plates were allowed to come to room temperature and ATP levels were measured after adding 1.5 μL of CellTiter-Glo and incubation for 10 min at room temperature, as a surrogate for cell viability.

Assay Development for PI3K inhibitor retest: All CCLs were plated in 384-well opaque, white assay plates at one column each of 250, 500, 1000, and 2000 cells/well in 30 μL/well of their specified medium. Assay plates were incubated for 4 days at 37°C/5% CO_2. After the incubation period was over, the plates were removed from the incubator and allowed to come to room temperature. Subsequently, 30 μL of a 1:2 dilution of CellTiter-Glo reagent in PBS was added to each well and the plates were

shaken at room temperature for a few seconds. The plates were then incubated at room temperature for at least 10 min and read out using luminescence 384-well settings.

PI3K inhibitor retest: To confirm the effect of the PI3K and AKT inhibitors on the *PIK3CA*-mutant CCLs, a retest was performed. CCLs were seeded in opaque, white 384 well plates at 2,000 cells/well and their optimized (Table 6.3) density in 30 μL/well of their specified medium (Suppl. Table S6.1 (digital)). Cells were incubated for 24 h at 37°C/5% CO_2 and the compounds were pinned at their respective concentrations with 100 nL/well. The plates were subsequently incubated at 37°C/5% CO_2 for 72 hours. After the incubation period was over, the plates were removed from the incubator and allowed to come to room temperature. Subsequently, 30 μL of a 1:2 dilution of CellTiter-Glo reagent in PBS was added to each well and the plates were shaken at room temperature for a few seconds. The plates were then incubated at room temperature for at least 10 min and read out using luminescence 384-well settings.

Data Analysis

Data processing: As described in Chapter 5, using 3- or 4-parameter non-linear regressions to a sigmoid function (MATLAB) for fitting concentration-response curves of percent-viability scores.

Several steps for quality control for concentration points/curves were employed before final AUC calculation: 1) if compound-transfer events were reported as failed by the instrument, the respective data was removed; 2) plate data were reviewed visually in GeneData and wells or whole plates that had faulty readouts were masked; 3) during curve-fitting, a QC method based on Cook's distance was employed and outlier points

were eliminated; 4) at least 5 concentration-data points were required to fit a curve.

Genetic Data: as described in Chapter 5. TCGA data was downloaded from the Broad Institute's Cancer Genome Analysis (CGA) group (http://www.broadinstitute.org/cancer/cga). All available data on 10/31/2013 was downloaded, which included samples from the following lineages: acute myeloid leukemia (197 samples), bladder urothelial carcinoma (28 samples), breast invasive carcinoma (776 samples), cervical squamous cell carcinoma and endocervical adenocarcinoma (39 samples), colon adenocarcinoma (155 samples), glioblastoma multiforme (291 samples), head and neck squamous cell carcinoma (306 samples), kidney renal clear cell carcinoma (417 samples), kidney renal papillary cell carcinoma (112 samples), lung adenocarcinoma (229 samples), lung squamous cell carcinoma (178 samples), ovarian serous cystadenocarcinoma (316 samples), pancreatic adenocarcinoma (57 samples), prostate adenocarcinoma (83 samples), rectum adenocarcinoma (69 samples), skin cutaneous melanoma (266 samples), stomach adenocarcinoma (221 samples), thyroid carcinoma (405 samples), uterine corpus eendometrial carcinoma (248 samples), unspecified lineage (5 samples)

Overlap calculations: All available TCGA datasets from the Broad Institute's Cancer Genome Analysis (CGA) group using the firehose tool (www.broadinstitute.org/cancer/cga/firehose) as well as the mutational data of the CCLE (www.broadinstitute.org/ccle) oncomap and hybrid capture datasets were used.[1,9] All annotated protein information was removed. hg18 data was transferred to hg19 data using the University of Santa Cruz's LiftOver Tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver).[10-12] All mutations were re-annotated into protein changes with the

chromosome, start and end positions, mutated allele and reference allele for all samples using the Oncotator Tool (www.broadinstitute.org/cancer/cga/oncotator) from the Broad Institute's CGA group.[13] Of note, some location positions and protein changes were annotated conflictingly using this tool. This was done consistently however, and therefore did not influence later analyses. The errors were manually corrected in the final data file to reflect the position as given by the "canonical" sequence on UniProt (Suppl. Table S6.2 (digital)). The TCGA data was filtered to only include the 1,667 genes also contained in the CCLE files. Overlap of mutations was determined by protein change and by amino acid location.

Enrichment analysis recurrent genes: p-values that quantify the enrichment of genetic alterations relative to ranked sensitivities measured for a single compound across many CCLs were calculated using a sorting-based enrichment-scoring algorithm.[57] The p-value for each compound and each genetic feature corresponds to the likelihood of seeing that pattern of alterations (or a stronger one) enriched among the ranked sensitivities by chance. To control for multiple hypothesis testing, the Benjamini-Hochberg procedure[58] was applied. This generated q-values (adjusted squared maximum p-values) and a false-discovery rate (FDR) cutoff $q < 0.1$ was applied for Suppl. Table S6.2 (digital).

Enrichment analysis fusion genes: To determine enrichments for the fusion genes, all AUC values were split into several sets of sensitive and non-responsive CCLs using cutoffs at the AUC value of each mutant CCL for each compound. P-values were calculated using Fisher exact test for the separation between the sensitive and the non-sensitive CCL sets at each AUC cutoff. The lowest p-value calculated for each

compound was determined to be the p-value of enrichment and the corresponding AUC cutoff was noted. A cutoff of $p < 0.05$ was applied for Suppl. Table S6.3 (digital).

References

(1) Barretina, J. *et al.* "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity." *Nature* **2012**, *483*, 603-607.

(2) Greenman, C., Wooster, R., Futreal, P. A., Stratton, M. R. & Easton, D. F. "Statistical analysis of pathogenicity of somatic mutations in cancer." *Genetics* **2006**, *173*, 2187-2198.

(3) Torkamani, A., Verkhivker, G. & Schork, N. J. "Cancer driver mutations in protein kinase genes." *Cancer Lett. (N.Y., NY, U.S.)* **2009**, *281*, 117-127.

(4) Domcke, S., Sinha, R., Levine, D. A., Sander, C. & Schultz, N. "Evaluating cell lines as tumour models by comparison of genomic profiles." *Nat. Commun.* **2013**, *4*, 2126.

(5) Ertel, A., Verghese, A., Byers, S. W., Ochs, M. & Tozeren, A. "Pathway-specific differences between tumor cell lines and normal and tumor tissue cells." *Mol. Cancer* **2006**, *5*, 55.

(6) Gillet, J. P. *et al.* "Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance." *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 18708-18713.

(7) Sandberg, R. & Ernberg, I. "Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI)." *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2052-2057.

(8) The Cancer Genome Atlas Research Network "Comprehensive genomic characterization defines human glioblastoma genes and core pathways." *Nature* **2008**, *455*, 1061-1068.

(9) Broad Institute Cancer Genome Analysis Group. *Firehose*, <http://www.broadinstitute.org/cancer/cga/firehose> (accessed: 9/20/2013).

(10) Kent, W. J. *et al.* "The human genome browser at UCSC." *Genome Res.* **2002**, *12*, 996-1006.

(11) Karolchik, D. *et al.* "The UCSC Genome Browser database: 2014 update." *Nucleic Acids Res.* **2014**, *42*, D764-770.

(12) University of California Santa Cruz Genome Bioinformatics Group. *Lift Genome Annotations*, <http://genome.ucsc.edu/cgi-bin/hgLiftOver> (accessed: 10/29/2013).

(13) Broad Institute Cancer Genome Analysis Group. *Oncotator*, <http://www.broadinstitute.org/cancer/cga/oncotator> (accessed: 10/30/2013).

(14) Reva, B., Antipin, Y. & Sander, C. "Predicting the functional impact of protein mutations: application to cancer genomics." *Nucleic Acids Res.* **2011**, *39*, e118.

(15) Mwenifumbo, J. C. & Marra, M. A. "Cancer genome-sequencing study design." *Nat. Rev. Genet.* **2013**, *14*, 321-332.

(16) Albers, C. A. *et al.* "Dindel: accurate indel calls from short-read data." *Genome Res.* **2011**, *21*, 961-973.

(17) Mitelman, F., Johansson, B. & Mertens, F. "The impact of translocations and gene fusions on cancer causation." *Nat. Rev. Cancer* **2007**, *7*, 233-245.

(18) Forbes, S. A. *et al.* "The Catalogue of Somatic Mutations in Cancer (COSMIC)." *Curr. Protoc. Hum. Genet.* **2008**, *Chapter 10*, 10.11.11-10.11.26.

(19) Forbes, S. A. *et al.* "COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer." *Nucleic Acids Res.* **2011**, *39*, D945-950.

(20) Bollag, G. *et al.* "Vemurafenib: the first drug approved for BRAF-mutant cancer." *Nature reviews. Drug discovery* **2012**, *11*, 873-886.

(21) Huynh, H., Soo, K. C., Chow, P. K. & Tran, E. "Targeted inhibition of the extracellular signal-regulated kinase kinase pathway with AZD6244 (ARRY-

142886) in the treatment of hepatocellular carcinoma." *Mol. Cancer Ther.* **2007**, *6*, 138-146.

(22) Ohren, J. F. *et al.* "Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition." *Nat. Struct. Mol. Biol.* **2004**, *11*, 1192-1197.

(23) Ascierto, P. A. *et al.* "MEK162 for patients with advanced melanoma harbouring NRAS or Val600 BRAF mutations: a non-randomised, open-label phase 2 study." *Lancet Oncol.* **2013**, *14*, 249-256.

(24) Jänne, P. A. *et al.* "Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study." *Lancet Oncol.* **2013**, *14*, 38-47.

(25) Rosell, R. *et al.* "Mutations in the tyrosine kinase domain of the EGFR gene associated with gefitinib response in non-small-cell lung cancer." *Lung Cancer* **2005**, *50*, 25-33.

(26) Mitsudomi, T. *et al.* "Mutations of the epidermal growth factor receptor gene predict prolonged survival after gefitinib treatment in patients with non-small-cell lung cancer with postoperative recurrence." *J. Clin. Oncol.* **2005**, *23*, 2513-2520.

(27) Brana, I. & Siu, L. L. "Clinical development of phosphatidylinositol 3-kinase inhibitors for cancer treatment." *BMC Med.* **2012**, *10*, 161.

(28) Nylander, S. *et al.* "Human target validation of phosphoinositide 3-kinase (PI3K)beta: effects on platelets and insulin sensitivity, using AZD6482 a novel PI3Kbeta inhibitor." *J. Thromb. Haemostasis* **2012**, *10*, 2127-2136.

(29) Ogino, S. *et al.* "Predictive and prognostic analysis of PIK3CA mutation in stage III colon cancer intergroup trial." *J. Natl. Cancer Inst.* **2013**, *105*, 1789-1798.

(30) Hirai, H. *et al.* "MK-2206, an allosteric Akt inhibitor, enhances antitumor efficacy by standard chemotherapeutic agents or molecular targeted drugs in vitro and in vivo." *Mol. Cancer Ther.* **2010**, *9*, 1956-1967.

(31) Luo, J., Manning, B. D. & Cantley, L. C. "Targeting the PI3K-Akt pathway in human cancer: rationale and promise." *Cancer Cell* **2003**, *4*, 257-262.

(32) Yuan, T. L. & Cantley, L. C. "PI3K pathway alterations in cancer: variations on a theme." *Oncogene* **2008**, *27*, 5497-5510.

(33) Parker, W. B. *et al.* "Effects of 2-chloro-9-(2-deoxy-2-fluoro-beta-D-arabinofuranosyl)adenine on K562 cellular metabolism and the inhibition of human ribonucleotide reductase and DNA polymerases by its 5'-triphosphate." *Cancer Res.* **1991**, *51*, 2386-2394.

(34) Bonate, P. L. *et al.* "Discovery and development of clofarabine: a nucleoside analogue for treating cancer." *Nat. Rev. Drug Discovery* **2006**, *5*, 855-863.

(35) Gould, K. L., Moreno, S., Tonks, N. K. & Nurse, P. "Complementation of the mitotic activator, p80cdc25, by a human protein-tyrosine phosphatase." *Science* **1990**, *250*, 1573-1576.

(36) Graf, C. *et al.* "Targeting ceramide metabolism with a potent and specific ceramide kinase inhibitor." *Mol. Pharmacol.* **2008**, *74*, 925-932.

(37) Gomez-Munoz, A. *et al.* "Short-chain ceramide-1-phosphates are novel stimulators of DNA synthesis and cell division: antagonism by cell-permeable ceramides." *Mol. Pharmacol.* **1995**, *47*, 833-839.

(38) Gomez-Munoz, A., Frago, L. M., Alvarez, L. & Varela-Nieto, I. "Stimulation of DNA synthesis by natural ceramide 1-phosphate." *Biochem. J.* **1997**, *325 (Pt 2)*, 435-440.

(39) Marks, K. M. *et al.* "The selectivity of austocystin D arises from cell line-specific drug activation by cytochrome P450 enzymes." *J. Nat. Prod.* **2011**, *74*, 567-573.

(40) Kipp, J. L. *et al.* "Gene expression profiling reveals Cyp26b1 to be an activin regulated gene involved in ovarian granulosa cell proliferation." *Endocrinology* **2011**, *152*, 303-312.

(41) Mukasa, C. *et al.* "Activin signaling through type IB activin receptor stimulates aromatase activity in the ovarian granulosa cell-like human granulosa (KGN) cells." *Endocrinology* **2003**, *144*, 1603-1611.

(42) Burger, M. T. *et al.* "Identification of NVP-BKM120 as a Potent, Selective, Orally Bioavailable Class I PI3 Kinase Inhibitor for Treating Cancer." *ACS Med. Chem. Lett.* **2011**, *2*, 774-779.

(43) Tanaka, H. *et al.* "The selective class I PI3K inhibitor CH5132799 targets human cancers harboring oncogenic PIK3CA mutations." *Clin. Cancer Res.* **2011**, *17*, 3272-3281.

(44) Blake, J. F. *et al.* "Discovery and preclinical pharmacology of a selective ATP-competitive Akt inhibitor (GDC-0068) for the treatment of human tumors." *J. Med. Chem.* **2012**, *55*, 8110-8127.

(45) Ross, R. L., Askham, J. M. & Knowles, M. A. "PIK3CA mutation spectrum in urothelial carcinoma reflects cell context-dependent signaling and phenotypic outputs." *Oncogene* **2013**, *32*, 768-776.

(46) Cong, L. *et al.* "Multiplex genome engineering using CRISPR/Cas systems." *Science* **2013**, *339*, 819-823.

(47) Mali, P. *et al.* "RNA-guided human genome engineering via Cas9." *Science* **2013**, *339*, 823-826.

(48) Galkin, A. V. *et al.* "Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK." *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 270-275.

(49) Christensen, J. G. *et al.* "Cytoreductive antitumor activity of PF-2341066, a novel inhibitor of anaplastic lymphoma kinase and c-Met, in experimental models of anaplastic large-cell lymphoma." *Mol. Cancer Ther.* **2007**, *6*, 3314-3322.

(50) Shaw, A. T. *et al.* "Crizotinib versus chemotherapy in advanced ALK-positive lung cancer." *N. Engl. J. Med.* **2013**, *368*, 2385-2394.

(51) Brasca, M. G. *et al.* "Optimization of 6,6-dimethyl pyrrolo[3,4-c]pyrazoles: Identification of PHA-793887, a potent CDK inhibitor suitable for intravenous dosing." *Bioorg. Med. Chem. Lett.* **2010**, *18*, 1844-1853.

(52) Bonvini, P. *et al.* "The effect of the cyclin-dependent kinase inhibitor flavopiridol on anaplastic large cell lymphoma cells and relationship with NPM-ALK kinase expression and activity." *Haematologica* **2009**, *94*, 944-955.

(53) Massard, C. *et al.* "A first in man, phase I dose-escalation study of PHA-793887, an inhibitor of multiple cyclin-dependent kinases (CDK2, 1 and 4) reveals unexpected hepatotoxicity in patients with solid tumors." *Cell Cycle* **2011**, *10*, 963-970.

(54) Bartram, C. R. *et al.* "Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia." *Nature* **1983**, *306*, 277-280.

(55) Puil, L. *et al.* "Bcr-Abl oncoproteins bind directly to activators of the Ras signalling pathway." *EMBO J.* **1994**, *13*, 764-773.

(56) Neshat, M. S., Raitano, A. B., Wang, H. G., Reed, J. C. & Sawyers, C. L. "The survival function of the Bcr-Abl oncogene is mediated by Bad-dependent and -independent pathways: roles for phosphatidylinositol 3-kinase and Raf." *Mol. Cell. Biol.* **2000**, *20*, 1179-1186.

(57) Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. *Introduction to algorithms*. 2nd edn, (MIT Press :McGraw-Hill, 2001).

(58) Benjamini, Y. & Hochberg, Y. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *J. R. Statist. Soc. B* **1995**, *57*, 289-300.

Page intentionally left blank.

Chapter Seven

Further outlook into small-molecule probe development for cancer

Summary of research presented

This thesis describes an effort to develop novel small-molecule probes and their application to research in cancer biology. Many small-molecule probes have been discovered by the Chemical Biology community. In order to use them as reliable tools, quality-control measures have been defined.[1,2] In Chapter Two, I describe the characterization of several small molecules previously annotated as probes, which were hits in a high-throughput screen for novel Hedgehog (Hh) pathway inhibitors. Some of these compounds most likely act on the Hh pathway through cross-signaling of their annotated targets. However, I showed that several of these compounds displace BODIPY-cyclopamine, a fluorescently labeled Smoothened (Smo) inhibitor, suggesting that they also interact with Smo. This additional activity should be considered when using them as probes to determine biological effects. For some of these compounds, adjusting the concentration can avoid the problem, but for others, additional probes might be necessary to confirm results.

The high-throughput screen not only presented known compounds but also provided a class of compounds with interesting stereochemistry-based structure–activity relationships (SAR) as inhibitors of the Hh pathway, as described in Chapter Three. We interpreted this striking stereochemical effect as indirect evidence for a selective interaction with cellular target(s). Further building-block based SAR generated two probes, BRD50837 and BRD9526, both of which inhibit the Hh pathway potently.

In Chapter Four, I investigated the mechanism of action of the two probes discovered in Chapter Three. They were tested in several epistasis and competition experiments, which showed that they acted similarly to the traditional pathway inhibitor

cyclopamine in some aspects, but different in others. These results are puzzling when interpreting the pathway using the traditional canonical model. However, the pathway has been shown to be very complex. Non-canonical signaling models are present and several small-molecule inhibitors target the same receptor, but cause different cellular outcomes.[3-5] This puzzling result is most likely due to the complexity of this enigmatic pathway. Identifying the compounds' target can realize their full potential as probes and enable further elucidation of the pathway. First studies to identify this target gave rise to a family of proteins, glutathione S-transferase μ (GSTMs) as potential interacting proteins.

In order to apply small-molecule probes directly to cancer research, a cancer cell-line (CCL) profiling effort is described in Chapter Five. Here, the response of 242 CCLs was assessed against an Informer Set of 355 compounds. Using publicly available genetic characterizations of the CCLs, novel hypotheses of cancer dependencies were generated with enrichment and clustering analyses. This approach detected already-known cancer dependencies, which gave us confidence in the process.

One important question arising from the data generated in Chapter Five and in an additional profiling effort by the Broad Institute's Chemical Biology Program and Platform was how to prioritize which hypothesis to pursue. For many hypotheses, little was known about the effects of enriched mutations or their biological relevance. Chapter Six describes a way to focus the enrichment on biologically relevant mutations. This new enrichment analysis generated hypotheses for specific mutations of several genes and showed that these enrichments would not have been found using all mutations in a certain gene. One hypothesis arising was that of differential sensitivity of

PIK3CA-mutant CCLs to PI3K and AKT inhibitors. In a retest of several *PIK3CA*-mutant CCLs, this differential sensitivity to PI3K inhibitors, but not to AKT inhibitors could be confirmed.

Future directions

In order to elucidate the Hh pathway further, identifying the target of BRD50837 and BRD9526 is necessary. A family of proteins, GSTMs, has been identified by biochemical affinity purification as interacting with the probes in cell lysates. More experiments are needed to determine the nature of the interaction and possible impacts on the Hh pathway. This could include biochemical assays to test inhibition of the enzyme *in vitro* as well as cellular assays to determine if the compounds' effect on cells is the same as with known inhibitors of GSTMs. Further experiments could include knock-down or overexpression of GSTMs and comparison of the resulting phenotype to that of small-molecule inhibition. Direct binding interactions could be studied using surface plasmon resonance or isothermal calorimetry. If the interaction of the probes with GSTMs is relevant to the Hh pathway, these experiments will give additional insight into Hh signaling.

The CCL profiling effort has generated many novel hypotheses of potential cancer dependencies. If these were to be confirmed, new ways of targeting cancer could be found for patients. One good experimental follow-up for this would be genetic editing using CRISPR.[6,7] We are currently working on proving differential sensitivity of *PIK3CA*-mutant CCLs to PI3K inhibitors by changing the mutational status of *PIK3CA*-mutant CCLs using genome editing. We are also attempting the same for *CDC25C*-

mutant CCLs and their sensitivity to clofarabine. If successful, retrospective analysis of data from clinical trials could be used to determine if patients show a similarly differential response and could alter the target group of the respective drugs.

Therefore, using small-molecule probes in research, our understanding of diseases can directly be altered. In some cases, new hypotheses can be generated that could rapidly translate into the clinic. In order to do so, the compounds must be reliable tools, and careful confirmation experiments have to be designed. However, the success of probe development and its influence on drug discovery show that the path is worthwhile to follow and many new discoveries can be made that ultimately change patients' lives.

References

(1) Oprea, T. I. *et al.* "A crowdsourcing evaluation of the NIH chemical probes." *Nat. Chem. Biol.* **2009**, *5*, 441-447.

(2) Bunnage, M. E., Chekler, E. L. & Jones, L. H. "Target validation using chemical probes." *Nat. Chem. Biol.* **2013**, *9*, 195-199.

(3) Jenkins, D. "Hedgehog signalling: emerging evidence for non-canonical pathways." *Cell. Signalling* **2009**, *21*, 1023-1034.

(4) Wang, Y., Zhou, Z., Walsh, C. T. & McMahon, A. P. "Selective translocation of intracellular Smoothened to the primary cilium in response to Hedgehog pathway modulation." *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2623-2628.

(5) Wang, Y. *et al.* "Selective identification of hedgehog pathway antagonists by direct analysis of smoothened ciliary translocation." *ACS Chem. Biol.* **2012**, *7*, 1040-1048.

(6) Cong, L. *et al.* "Multiplex genome engineering using CRISPR/Cas systems." *Science* **2013**, *339*, 819-823.

(7) Mali, P. *et al.* "RNA-guided human genome engineering via Cas9." *Science* **2013**, *339*, 823-826.

Appendix A

Supplementary Materials

Supplementary Figures

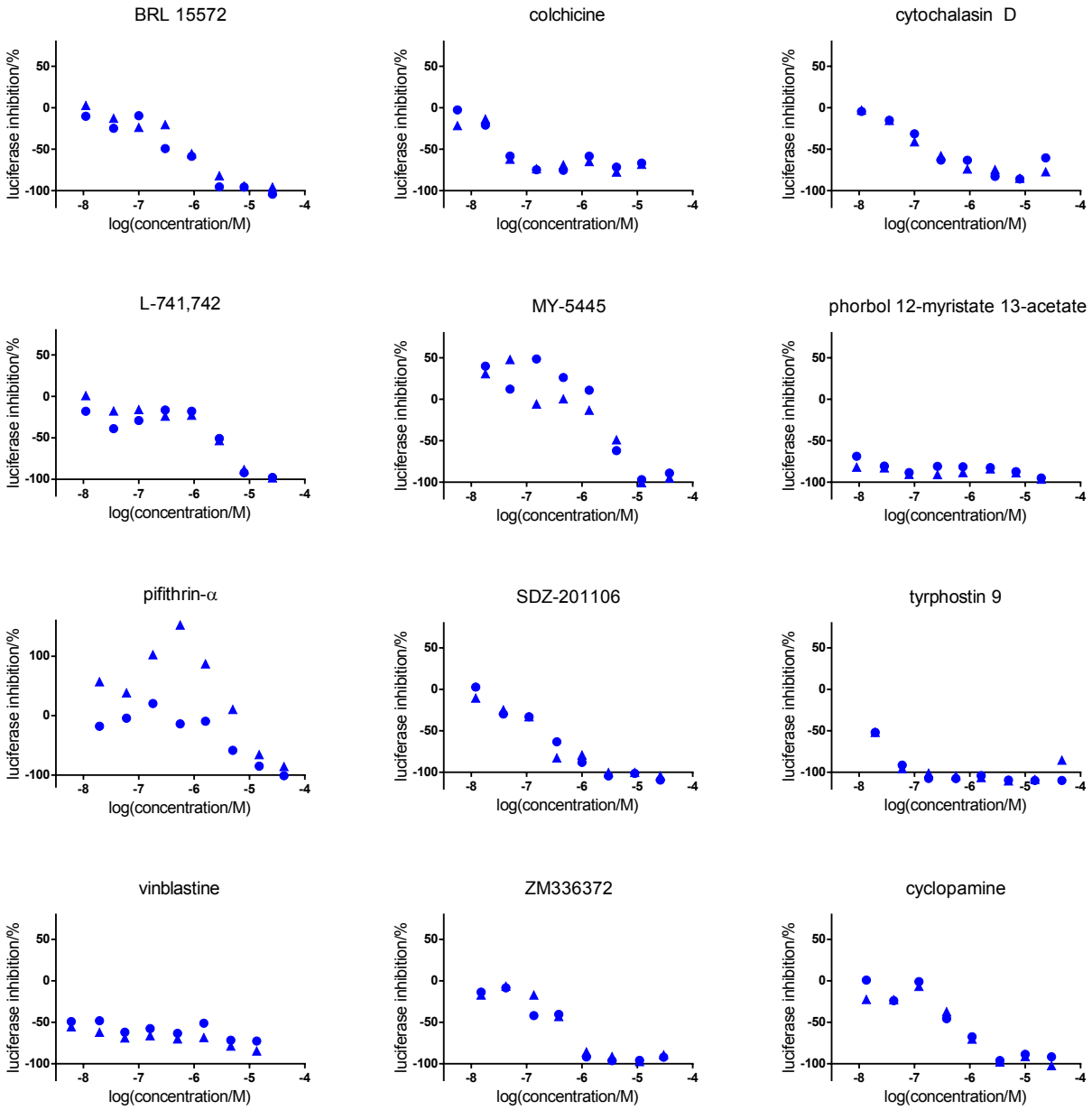


Figure S2.1. Inhibition of Shh-conditioned medium-induced Gli1 gene reporter luciferase activity treatment in Shh Light II cells with previously-annotated compounds and cyclopamine as a control after 30h. All values are shown and generated one experiment run in duplicate.

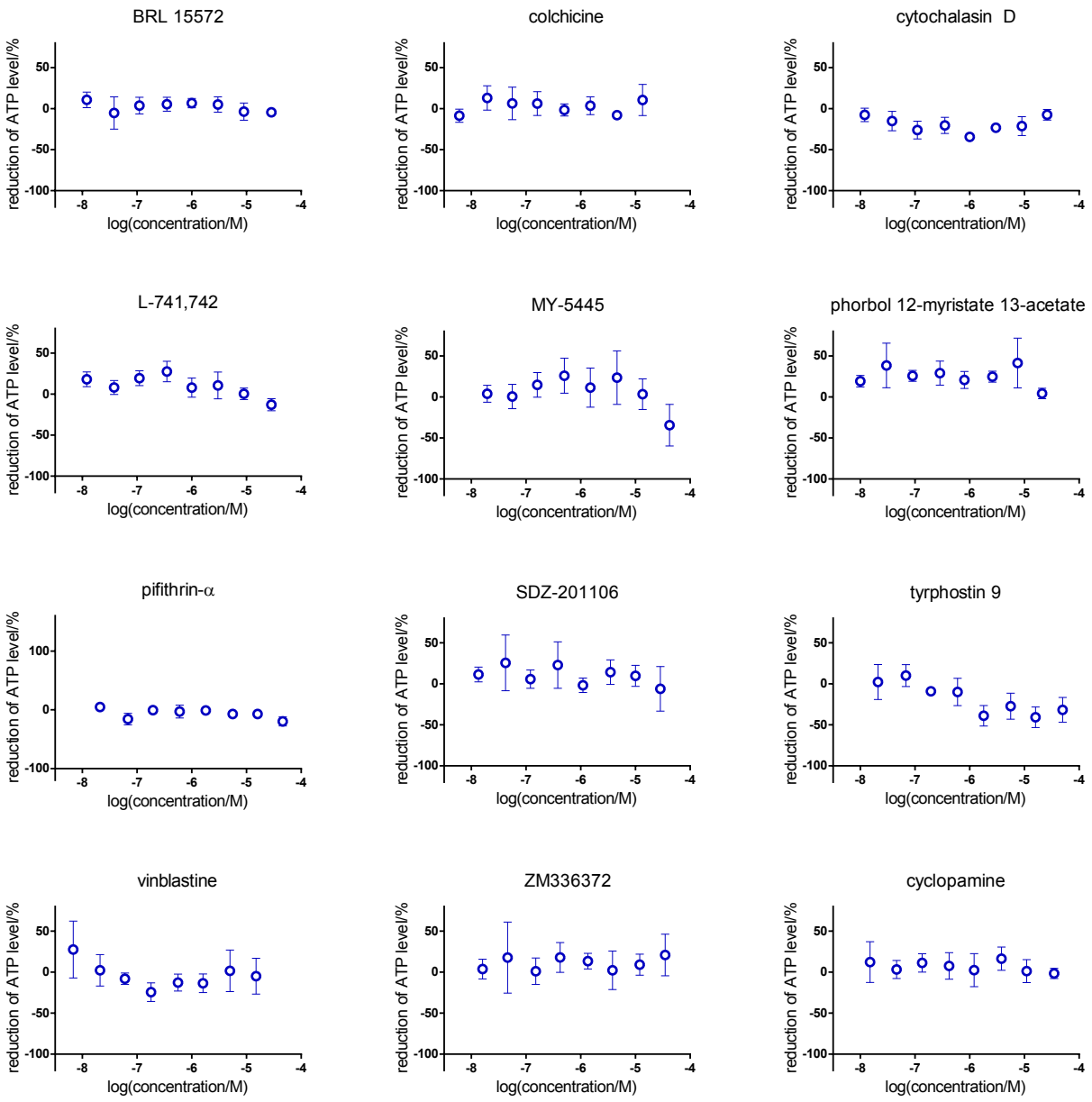


Figure S2.2. Viability of Shh Light II cells in response to 30 h treatment with previously annotated compounds and cyclopamine as a control. All values are shown and generated from two independent experiments run in duplicate (values are calculated average ± SD).

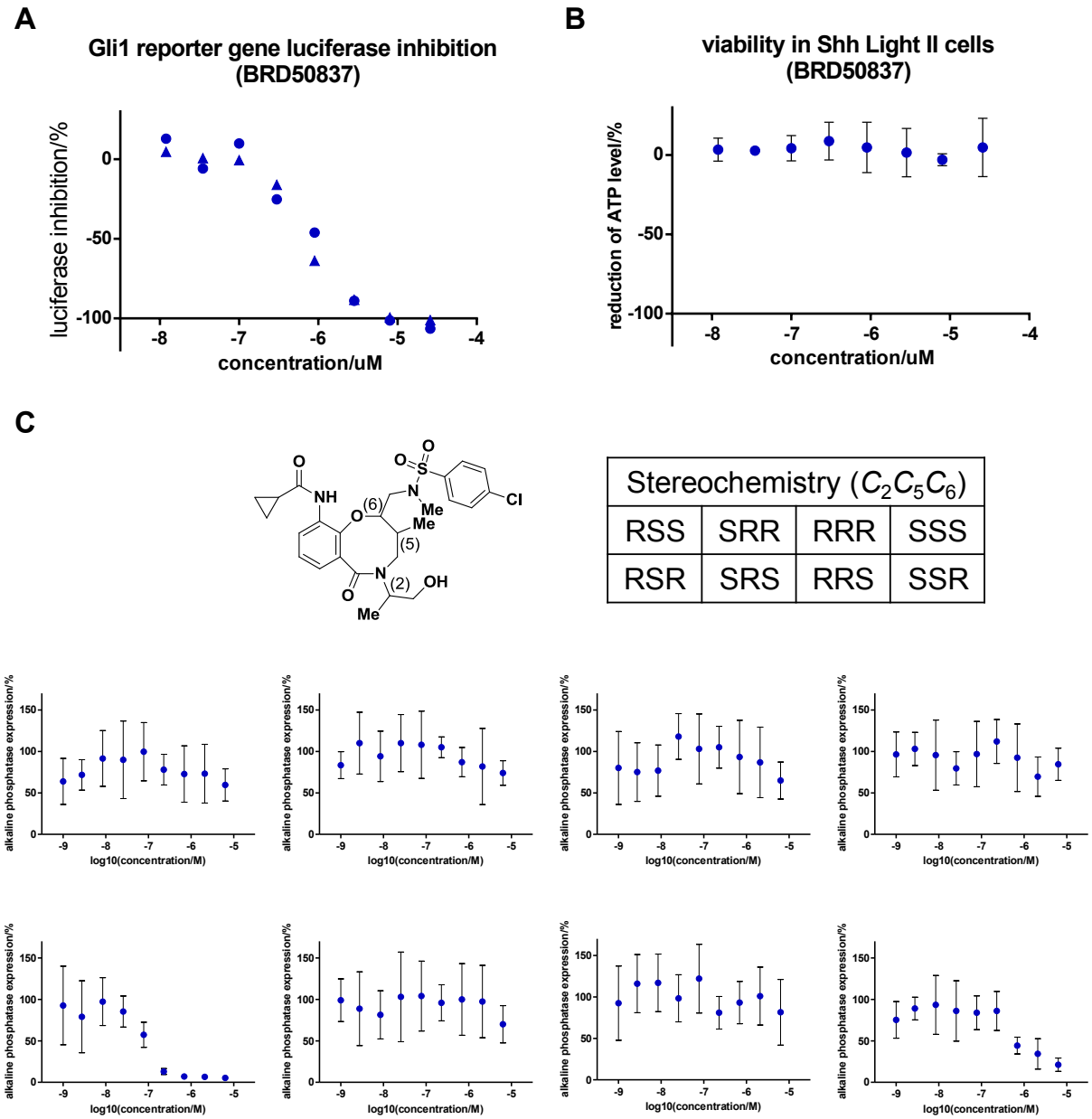


Figure S3.1. a) Dose response curve for inhibition of Gli1 gene reporter luciferase activity after 30 h treatment with BRD50837 in Shh Light II cells. All values are shown and generated from one independent experiment run in duplicate. b) Viability of Shh Light II cells in response to 30 h treatment with BRD50837. All values are shown and generated from two independent experiments run in duplicate (values are calculated average ± SD). c) Inhibition of Shh conditioned medium (CM) induced differentiation of C3H10T1/2 cells for all eight stereoisomers of BRD50837. The RSR and the SSR isomers are active with the RSR isomer being the more potent isomer. All values are shown and generated from three independent experiments run in duplicate (values are calculated average ± SD).

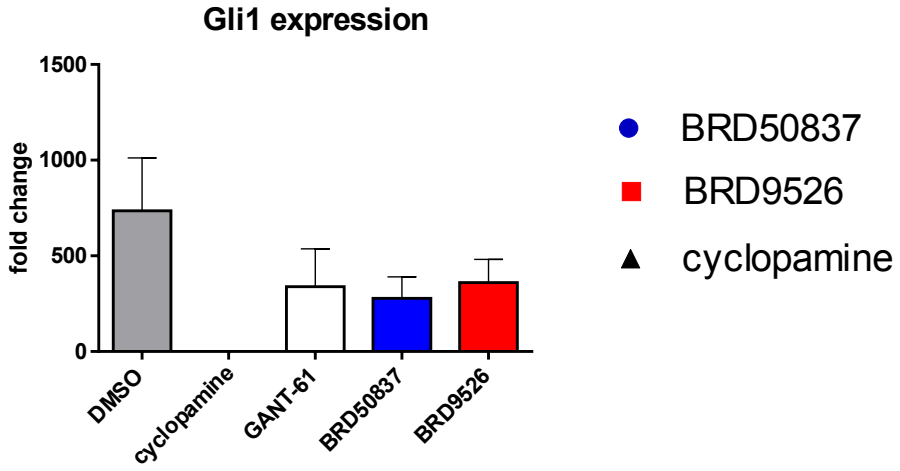


Figure S3.2. Gli expression for C3H10T1/2 cells induced with Shh-conditioned medium, treated for 48 hours with DMSO, 10 μM cyclopamine, 10 μM GANT-61, 1 μM BRD50837, or 1 μM BRD9526. All values are shown and generated from three independent experiments run in triplicate (c) (values are calculated average ± SD).

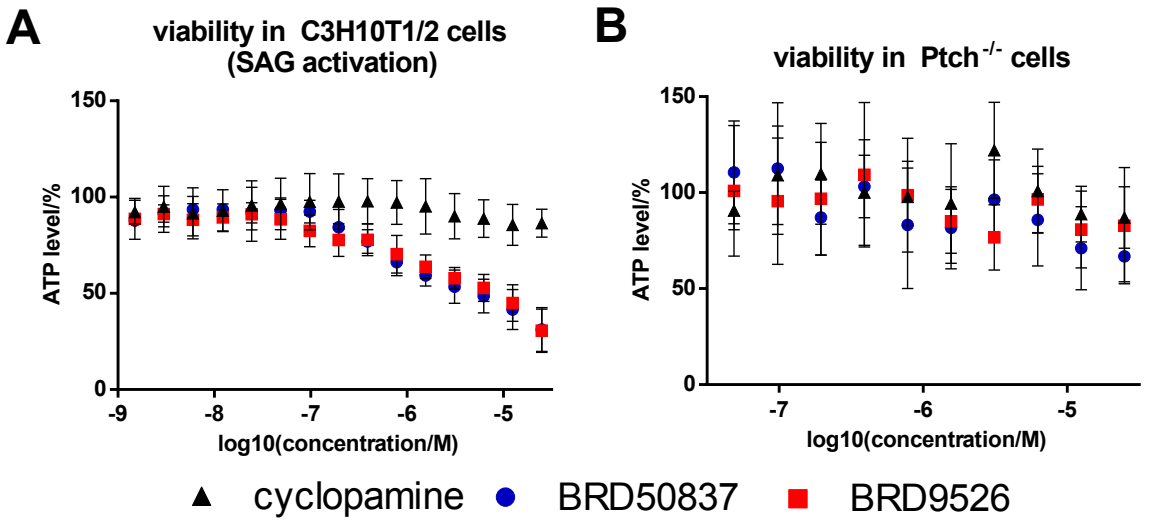


Figure S4.1. All values are shown and generated from three independent experiments run in duplicate (values are calculated average ± SD). a) Viability of SAG-induced C3H10T1/2 cells in response to 48 h treatment with BRD50837, BRD9526, or cyclopamine. b) Viability of Ptch$^{-/-}$ cells in response to 48 h treatment with BRD50837, BRD9526, or cyclopamine.

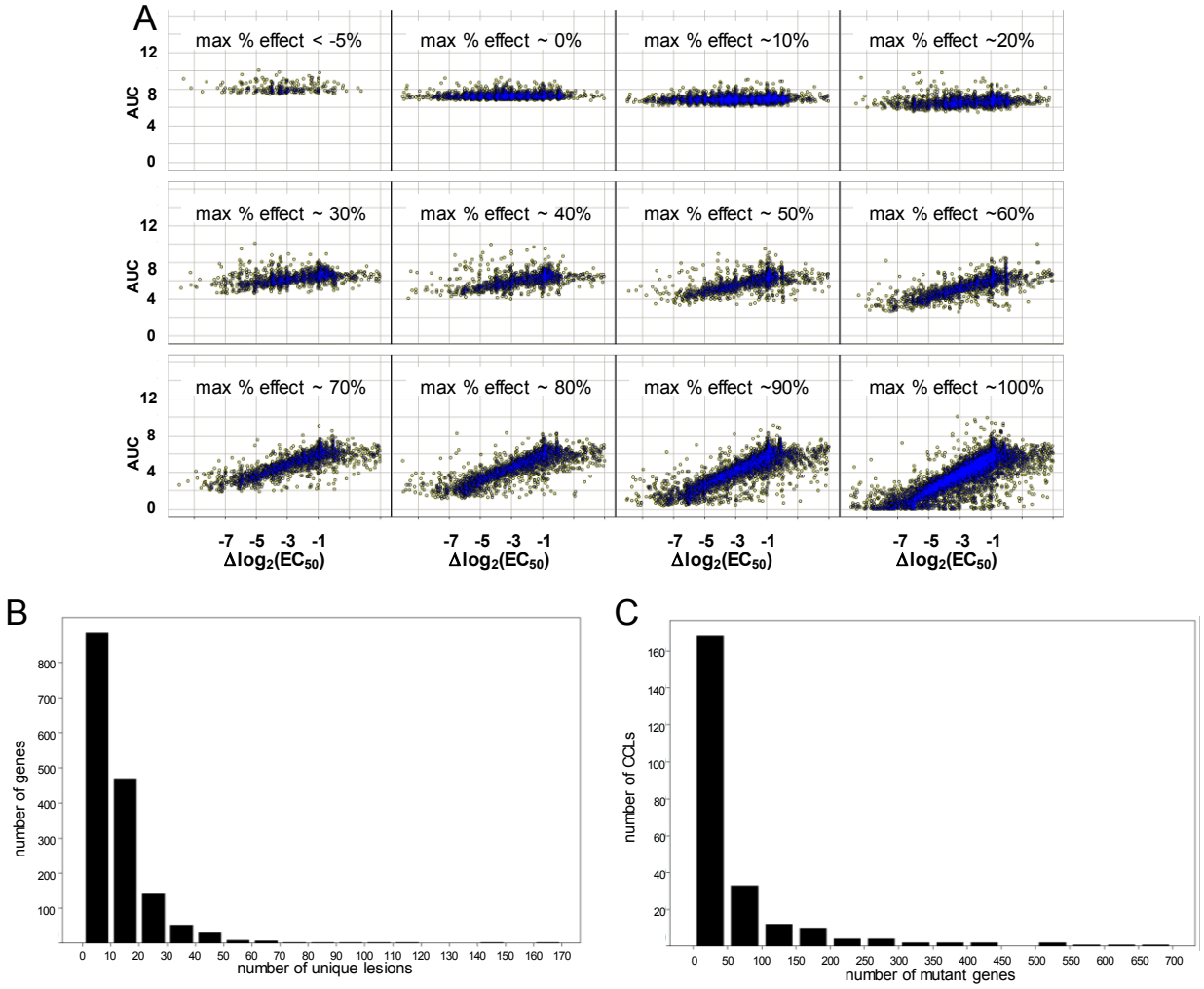


Figure S5.1. Properties of CCL profiling and genetic data. Area under concentration-response curves (AUC) accounts for both EC_{50} and strength of effect (**A**)**.** At low percent effect (i.e., when cell viability is relatively unaffected by compound treatment), AUC is essentially independent of relative EC_{50}. In contrast, as percent effect increases, the dependence of AUC on EC_{50} (as judged by the slope of their correlation) increases such that at 100% effect, changes in AUC are equivalent to changes in log(EC_{50}) (slope=1). Data presented represent a summary of 37,592 curve-fits (74.1% of all experiments in this study) for which the EC_{50} estimate was greater than 1/8 of the lowest concentration tested and less than 8X the highest concentration tested. Relative EC_{50}s were computed relative to the highest concentration tested for each compound, and strengths of effect were binned into groups centered on the indicated values for trellis display**.** Distributions of unique lesions (**B**) and frequencies of genes mutated in CCLs tested (**C**). The median CCL has mutations in 75 genes (5 percent of total genes sequenced). A large fraction of genes has several unique lesions.

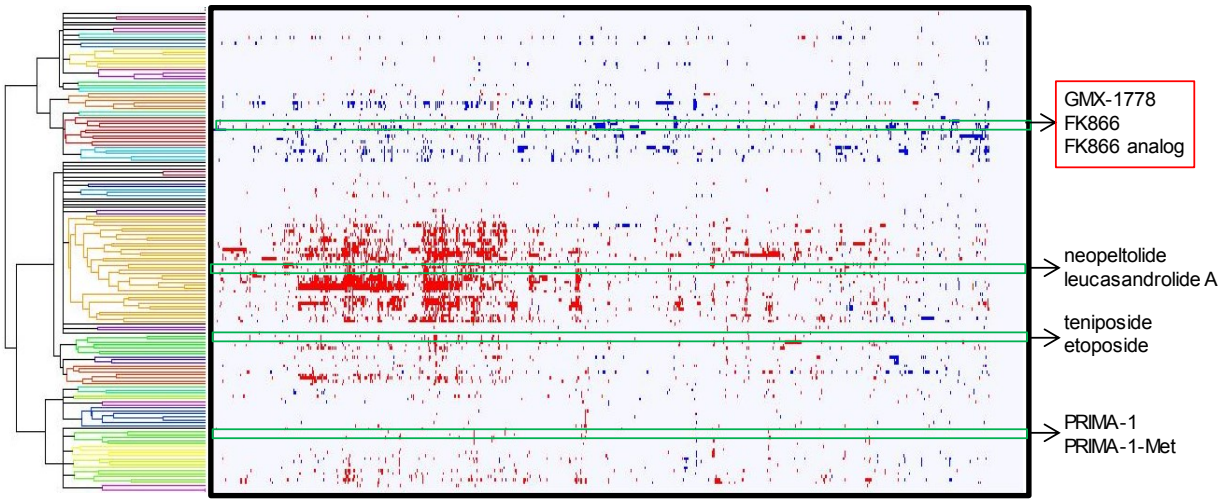


Figure S5.2. Properties of global connections. Dendrogram of all compounds used in the global analysis (using cosine distance in complete-linkage analysis); boxed cluster is described in the main text.

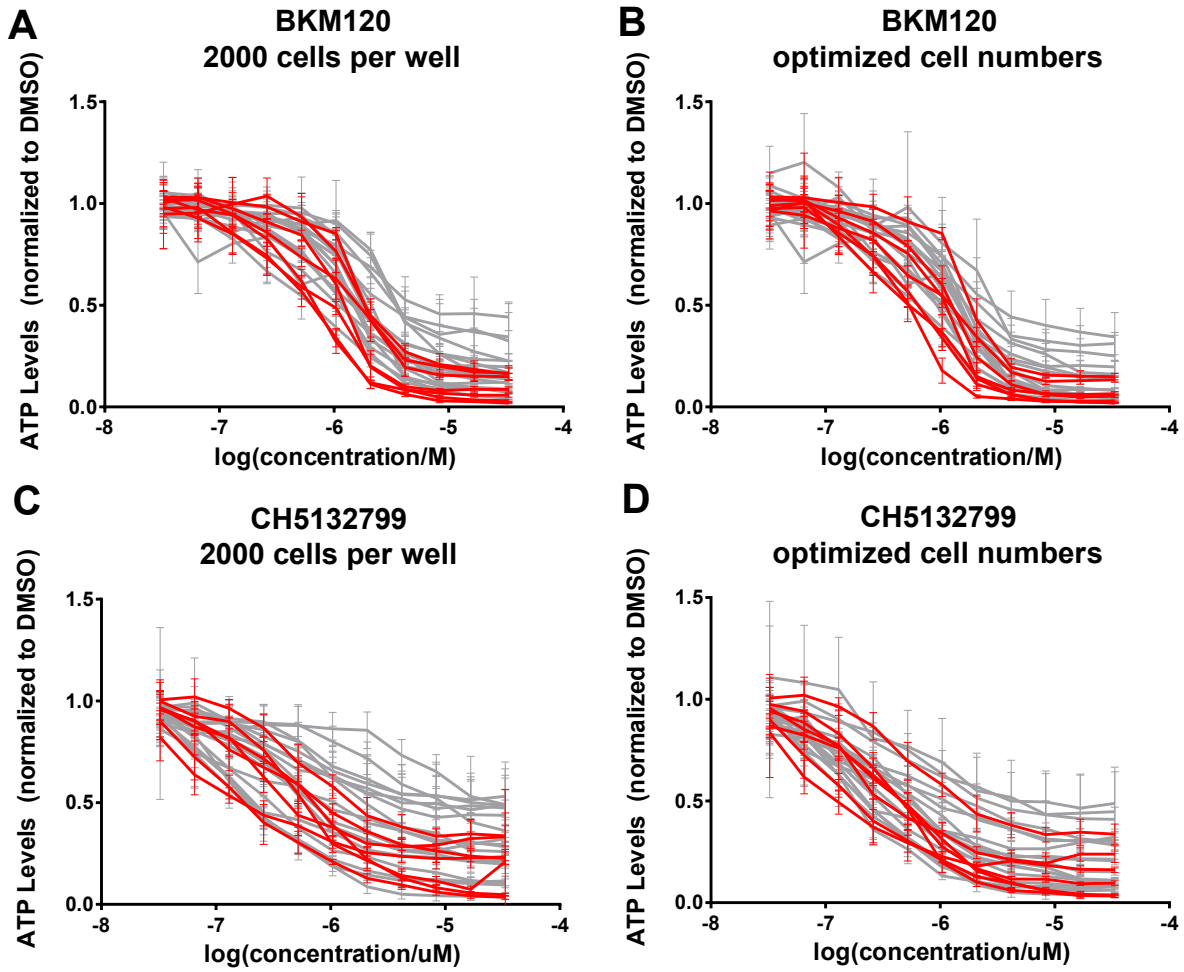


Figure S6.1. Dose-response curves of PI3K inhibitor re-test in 384-well format. All values are shown and were generated in two independent experiments run in triplicate (values are mean ± SD). Red dose-response curves are of *PIK3CA* E542K-mutant CCLs. a-b) dose-response curves for BKM120 in mutant *PIK3CA* CCLs at 2000 cells/well (a) or optimized densities (b). c-d) dose-response curves for CH5132799 in mutant *PIK3CA* CCLs at 2000 cells/well (c) or optimized densities (d).

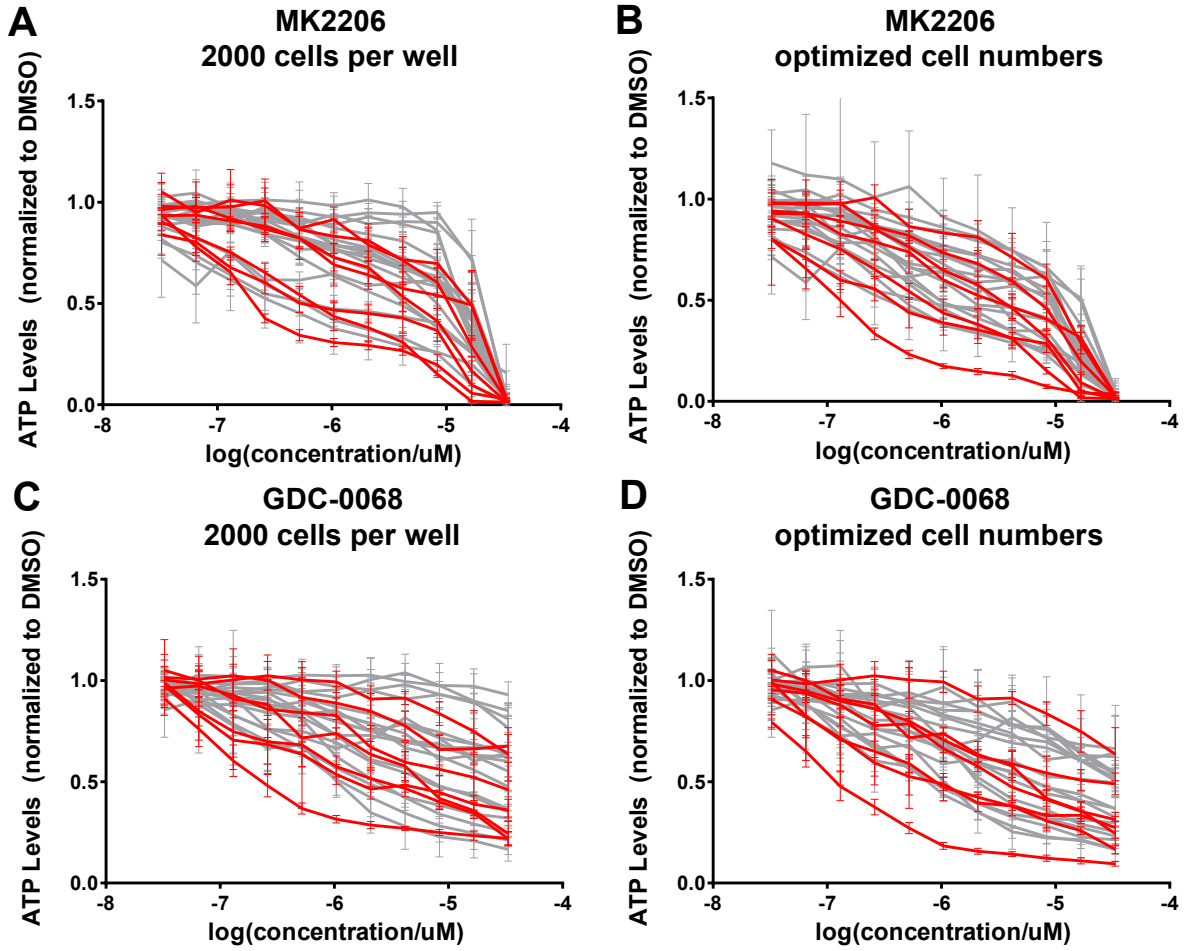
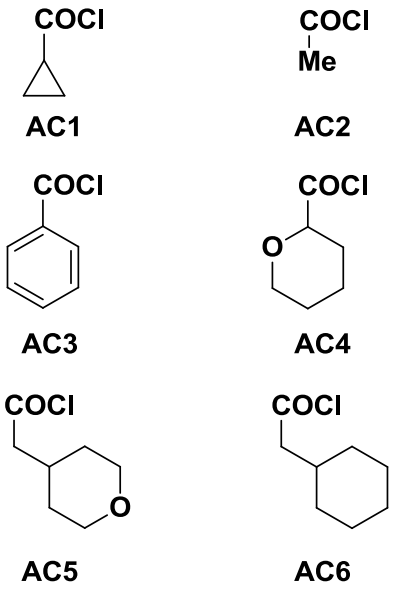


Figure S6.2. Dose-response curves of AKT inhibitor re-test in 384-well format. All values are shown and were generated in two independent experiments run in triplicate (values are mean ± SD). Red dose-response curves are of *PIK3CA* E542K-mutant CCLs. a-b) dose-response curves for MK2206 in mutant *PIK3CA* CCLs at 2000 cells/well (a) or optimized densities (b). c-d) dose-response curves for GDC-0068 in mutant *PIK3CA* CCLs at 2000 cells/well (c) or optimized densities (d).

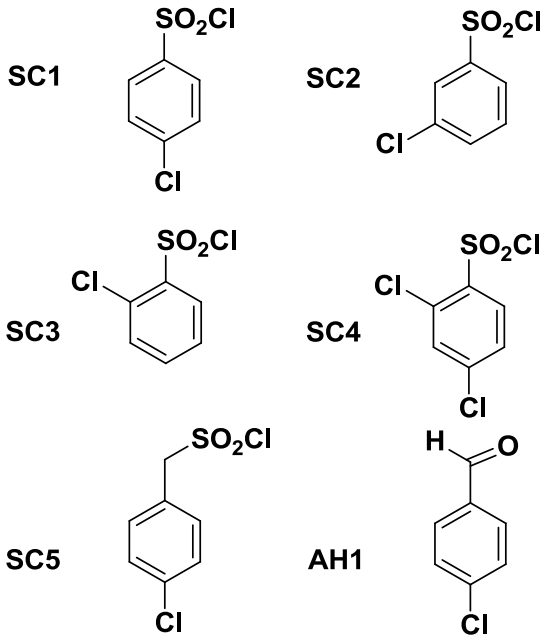
Supplementary Schemes

Scheme S3.1. Building blocks used for N-capping.

R_2

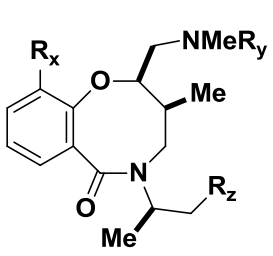


R_3

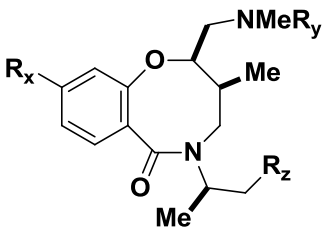


Supplementary Tables

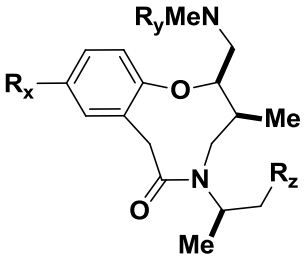
Table S3.1. Dose response curves of osteoblast differentiation and viability of C3H10T1/2 cells induced to differentiate by Hh conditioned medium (CM), EC_{50} of inhibition of osteoblast differentiation and viability, and PBS solubility of analogs of BRD50837



A



B



C

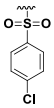
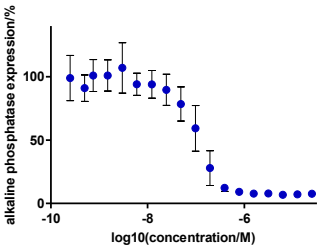
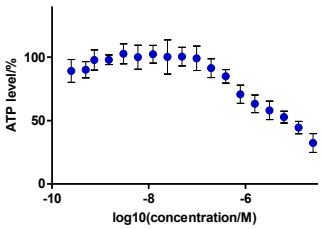
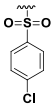
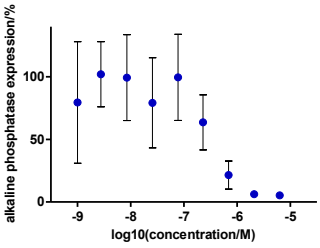
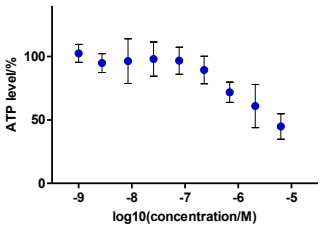
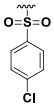
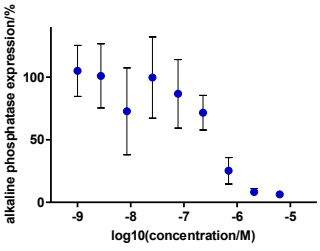
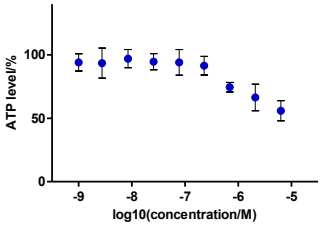
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
7 (BRD 50837)	A						0.09	6.41	64.3
8	A						0.22	3.78	-
9	A						0.29	b	-

Table S3.1. (Continued)

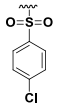
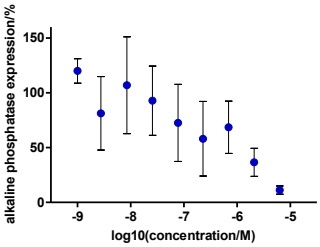
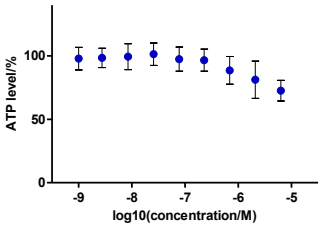
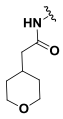
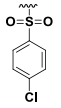
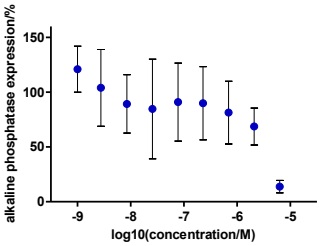
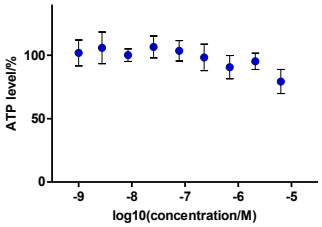
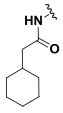
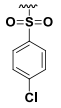
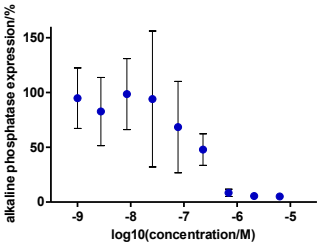
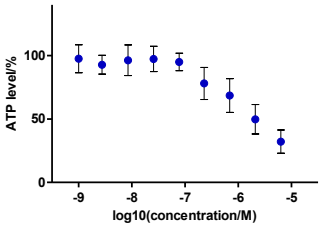
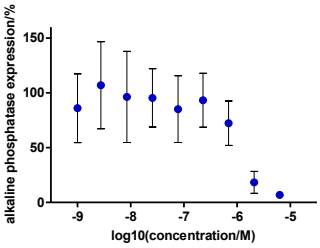
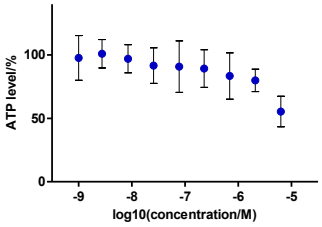
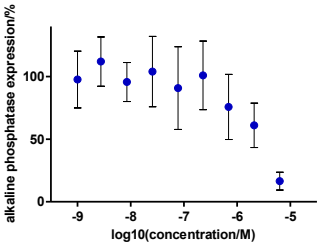
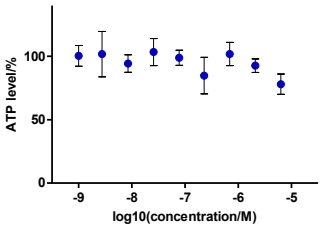
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
10	A						0.91	b	-
11	A						2.03	b	-
12	A						0.08	1.97	1.1
13	A						0.95	b	-
14	A						1.45	b	-

Table S3.1. (Continued)

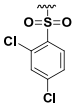
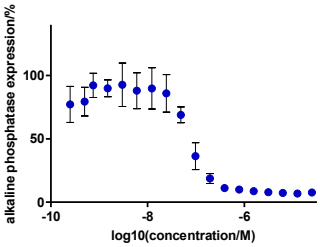
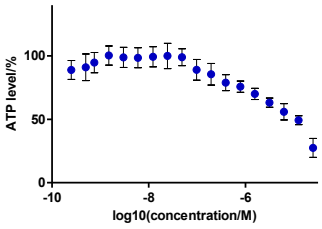
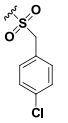
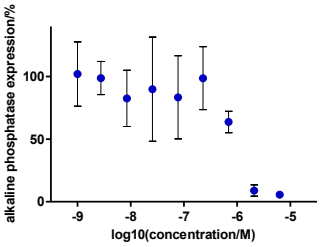
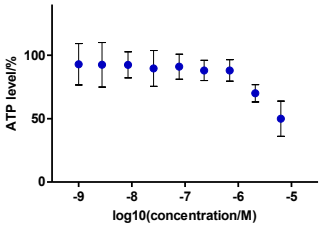
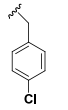
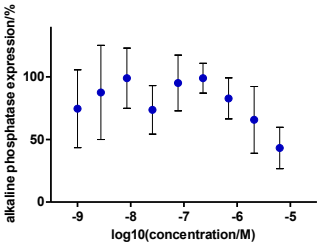
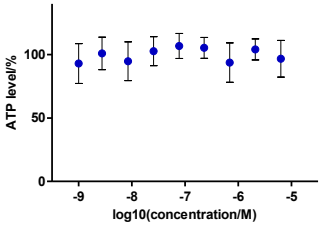
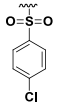
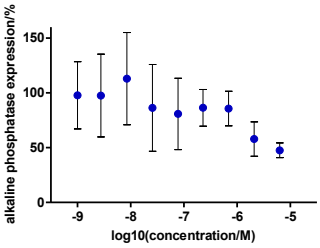
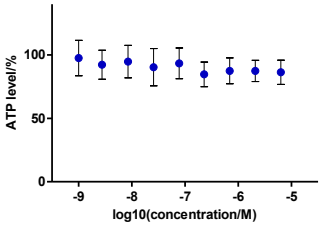
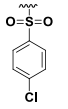
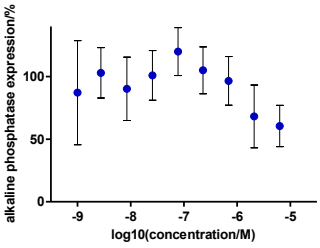
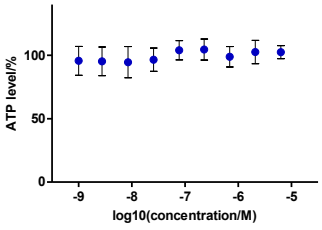
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
15 (BRD 9526)	A						0.06	7.84	57.4
16	A						0.76	5.96	-
17	A						3.37	b	-
18	A						4.03	b	-
19	A						b	b	-

Table S3.1. (Continued)

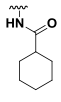
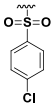
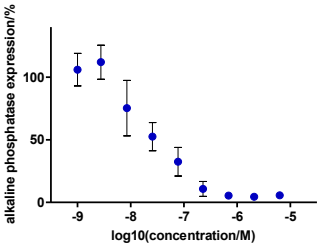
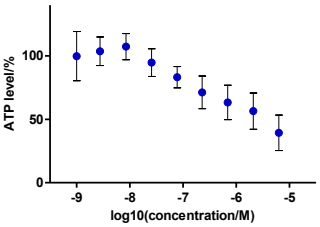
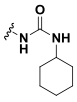
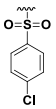
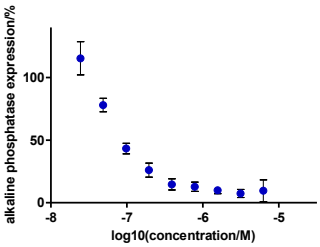
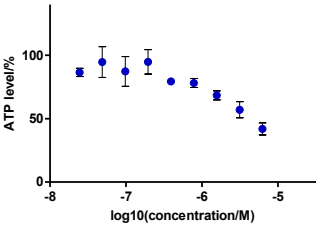
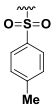
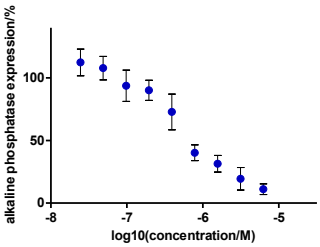
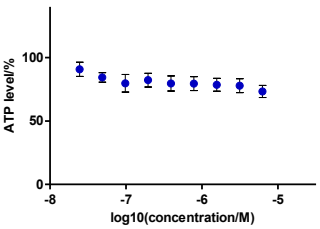
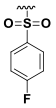
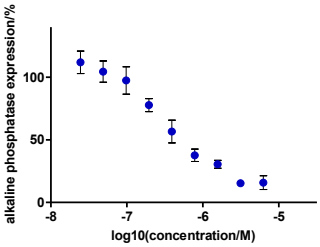
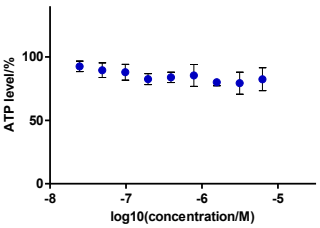
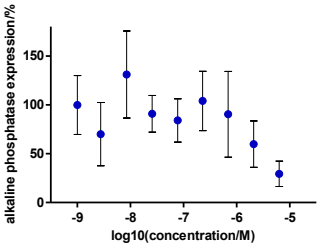
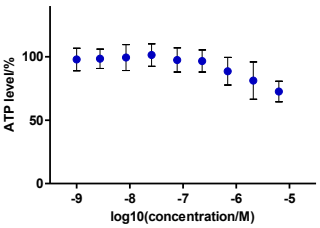
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
20	A						0.03	2.00	5.5
21	A						0.08	4.74	3.9
22	A						0.58	b	-
23	A						0.45	b	-
24	A						2.86	b	-

Table S3.1. (Continued)

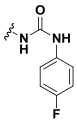
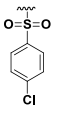
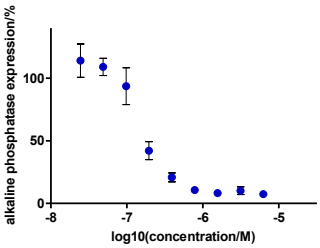
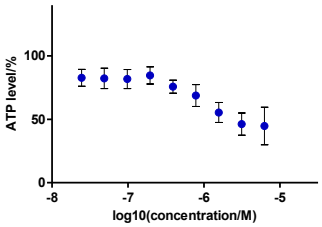
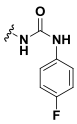
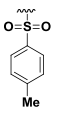
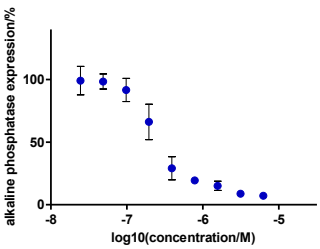
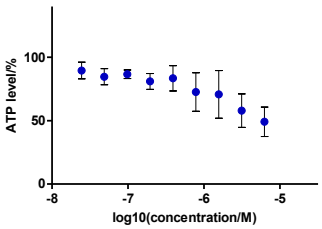
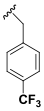
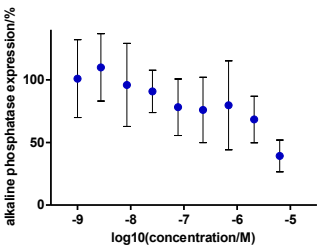
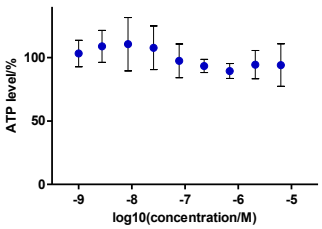
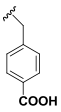
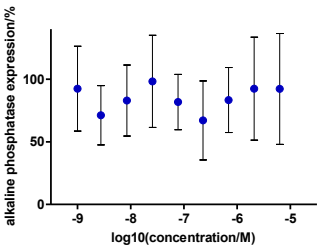
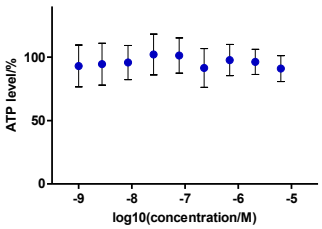
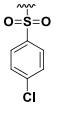
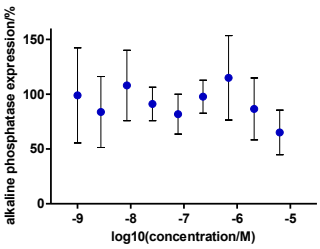
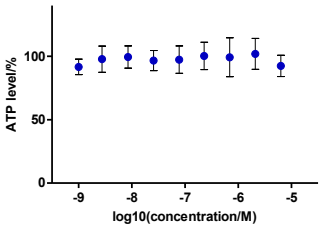
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
25	A						0.18	2.71	-
26	A						0.22	b	0.7
27	A						2.32	b	-
28	A						b	b	-
29	B						b	b	-

Table S3.1. (Continued)

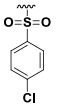
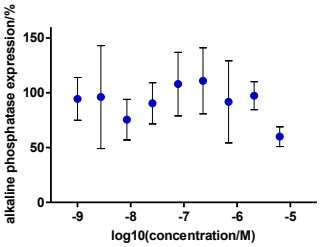
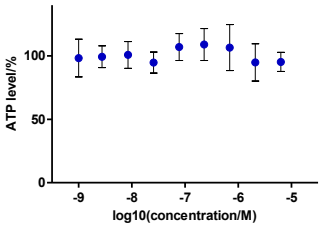
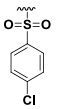
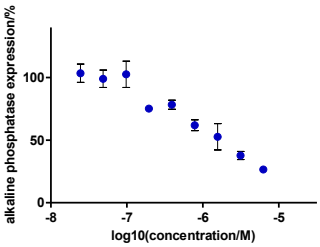
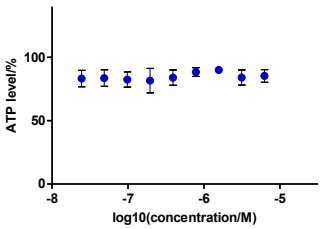
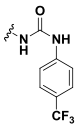
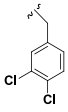
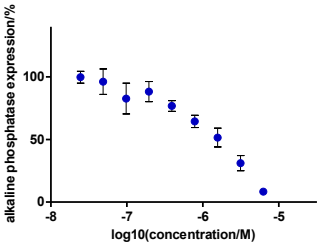
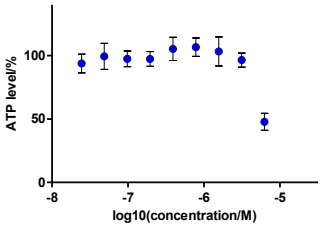
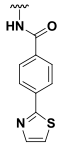
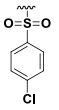
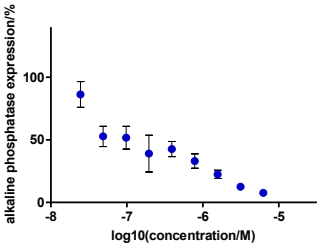
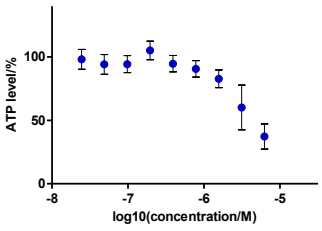
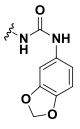
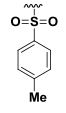
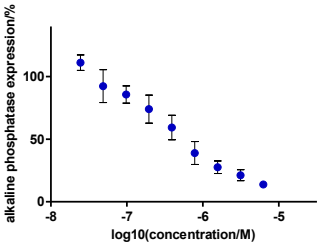
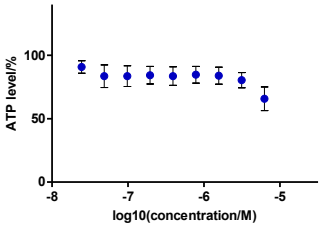
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
30	C						b	b	-
31	A						1.29	b	-
32	A						1.06	6.24	-
33	A						0.14	4.31	-
34	A						0.41	b	-

Table S3.1. (Continued)

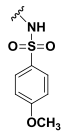
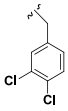
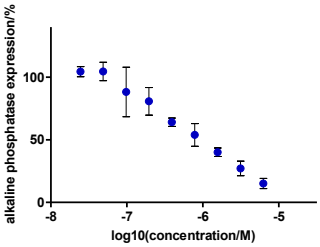
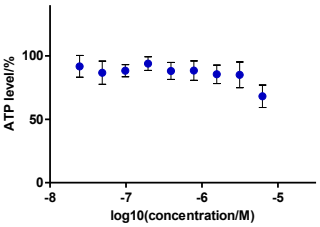
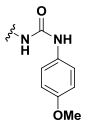
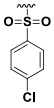
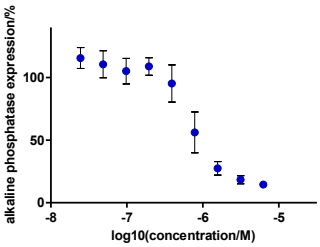
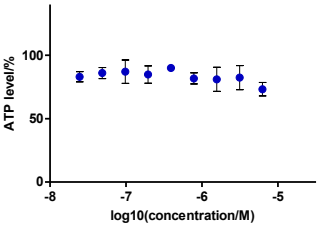
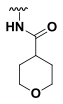
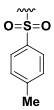
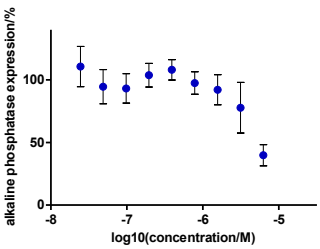
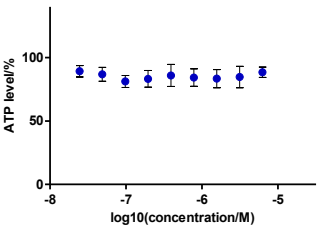
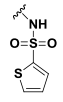
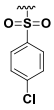
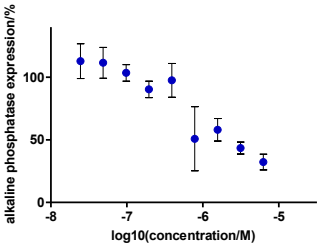
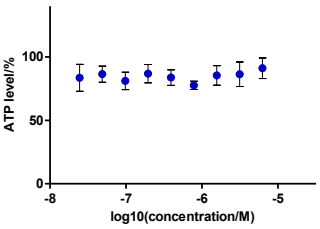
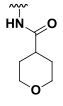
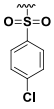
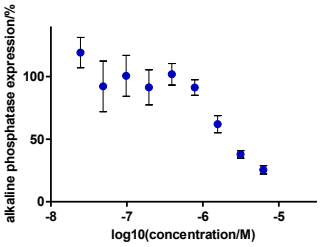
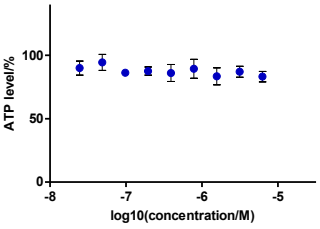
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
35	A						0.71	b	-
36	A						0.78	b	-
37	A						4.75	b	-
38	A						1.87	b	-
39	A						2.02	b	-

Table S3.1. (Continued)

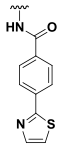
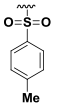
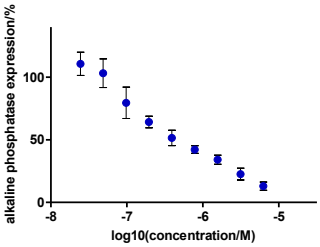
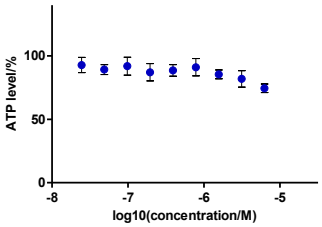
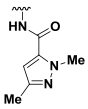
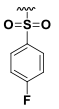
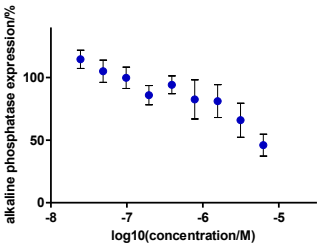
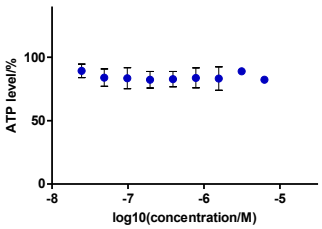
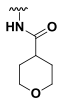
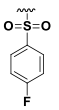
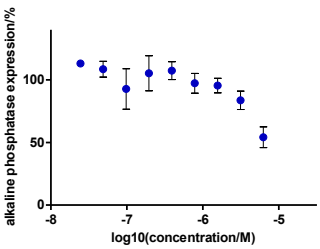
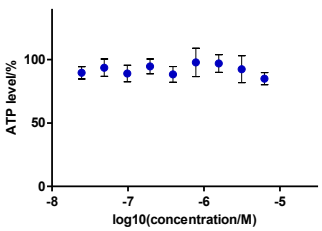
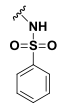
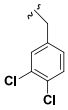
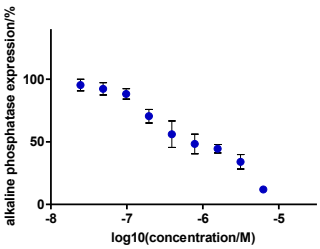
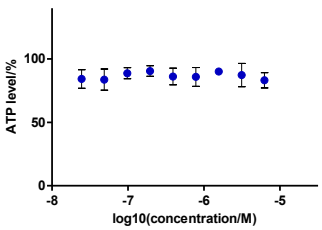
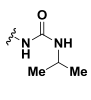
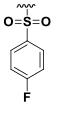
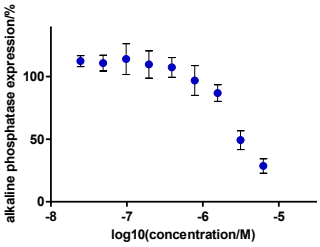
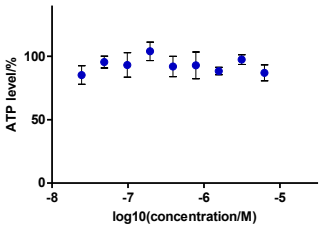
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
40	A						0.41	b	-
41	A						5.14	b	-
42	A						6.20	b	-
43	B						0.73	b	-
44	A						3.01	b	-

Table S3.1. (Continued)

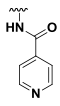
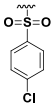
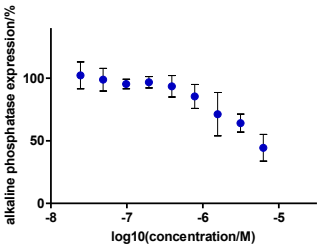
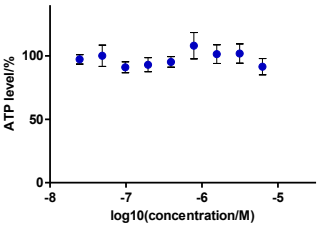
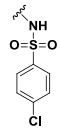
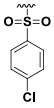
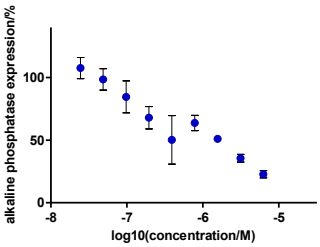
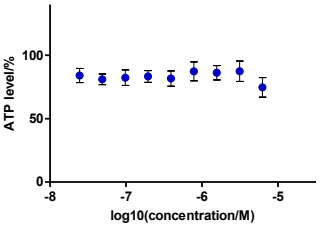
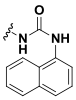
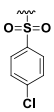
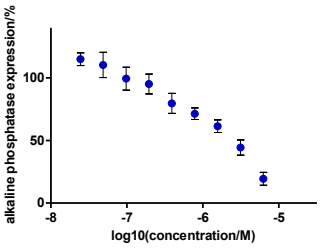
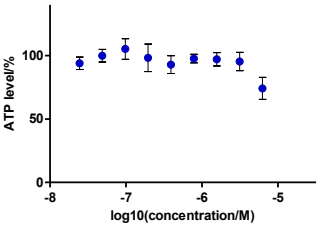
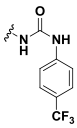
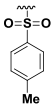
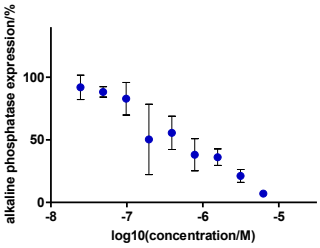
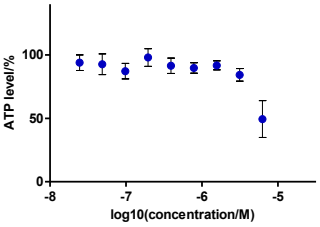
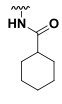
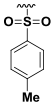
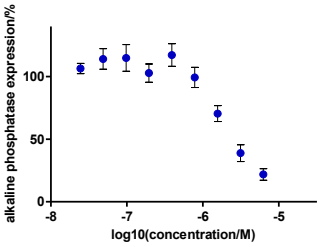
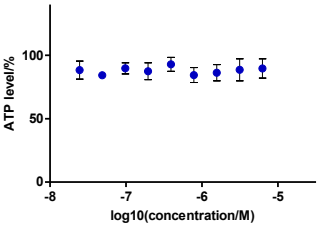
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
45	B						3.96	b	-
46	A						0.98	b	-
47	A						1.71	b	-
48	A						3.63	b	-
49	C						2.28	b	-

Table S3.1. (Continued)

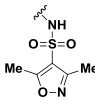
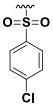
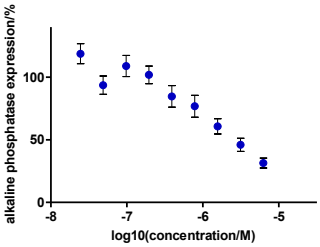
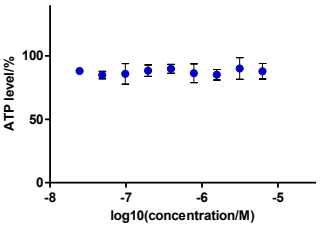
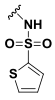
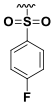
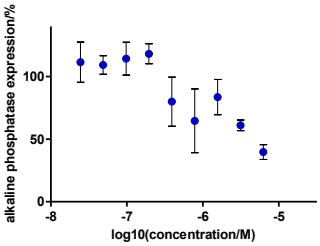
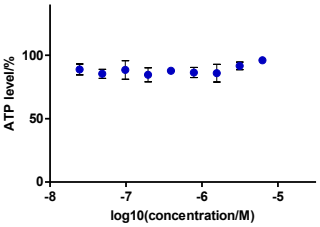
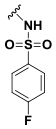
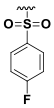
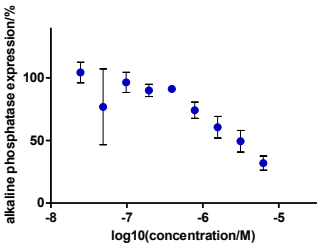
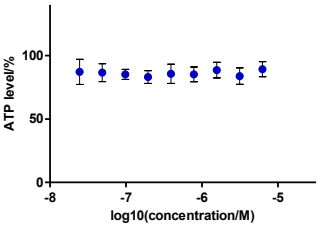
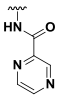
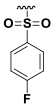
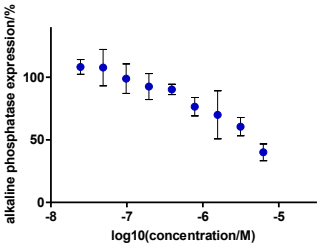
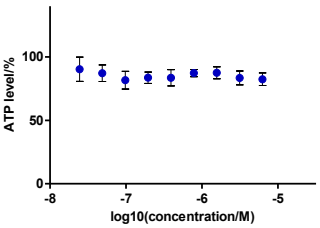
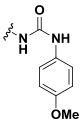
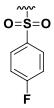
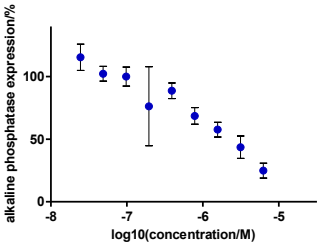
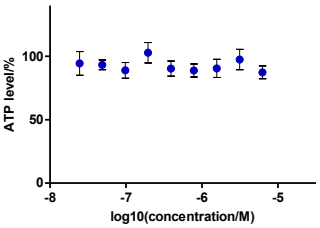
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
50	A						2.15	b	-
51	A						3.61	b	-
52	A						2.12	b	-
53	A						3.64	b	-
54	A						1.64	b	-

Table S3.1. (Continued)

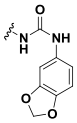
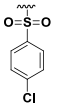
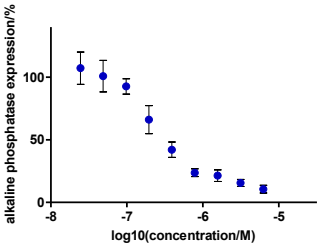
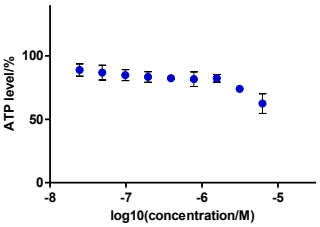
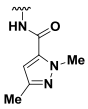
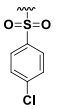
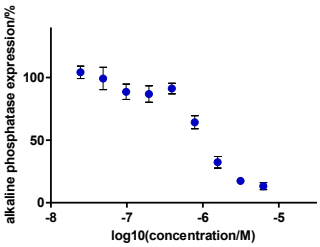
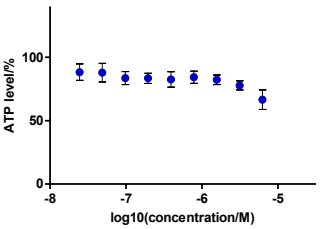
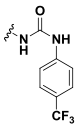
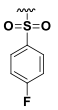
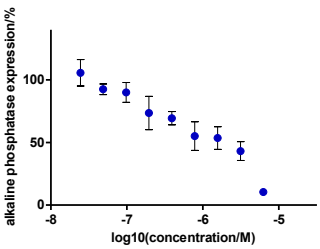
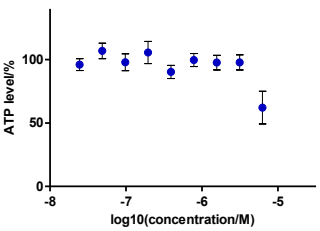
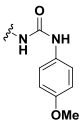
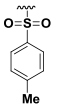
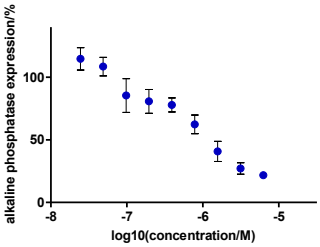
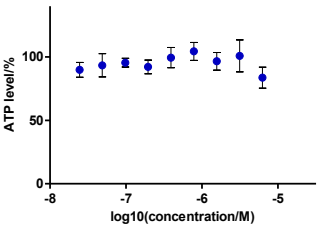
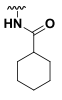
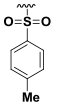
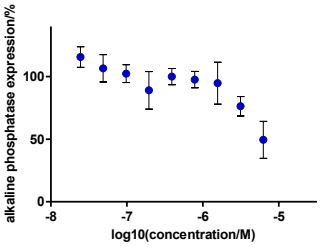
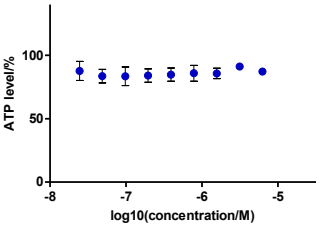
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
55	A						0.29	b	-
56	A						0.91	b	-
57	A						0.90	b	-
58	A						0.97	b	-
59	B						4.95	b	-

Table S3.1. (Continued)

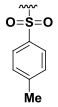
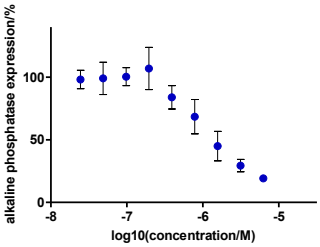
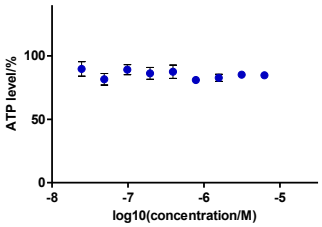
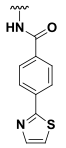
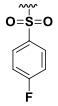
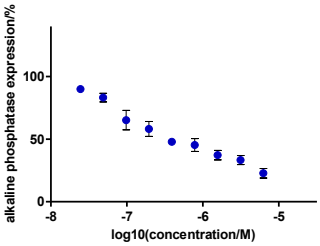
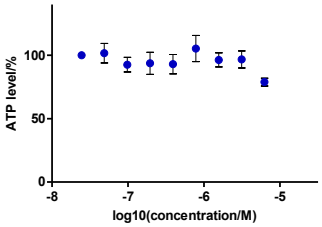
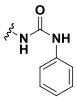
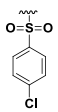
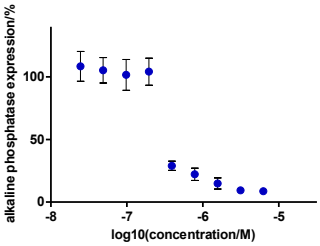
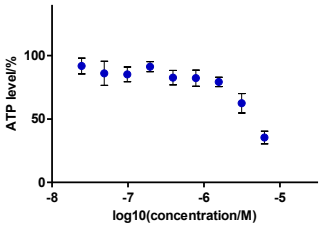
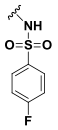
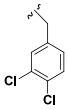
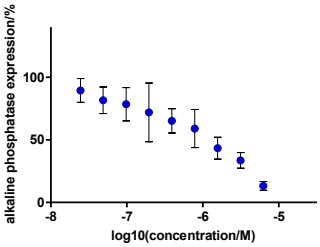
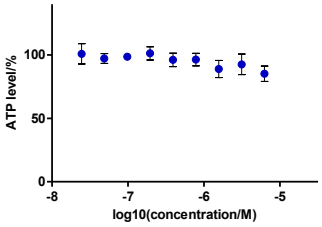
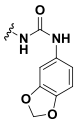
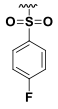
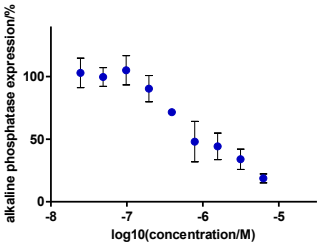
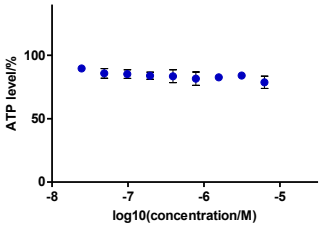
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
60	A						1.26	b	-
61	A						0.38	b	-
62	A						0.29	4.15	-
63	A						0.70	b	-
64	A						0.88	b	-

Table S3.1. (Continued)

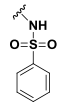
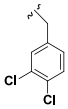
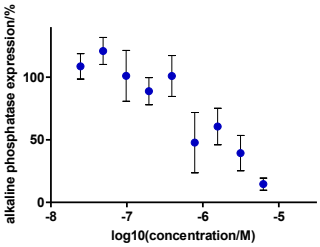
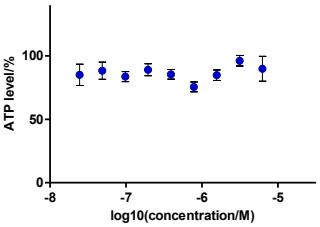
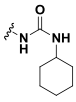
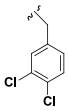
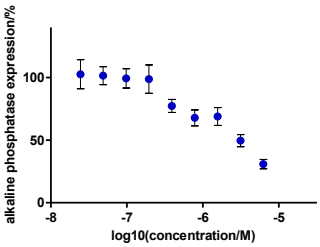
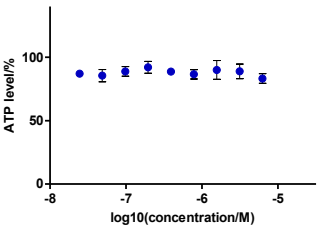
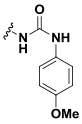
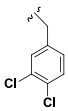
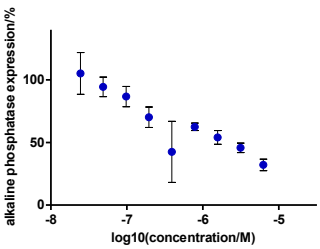
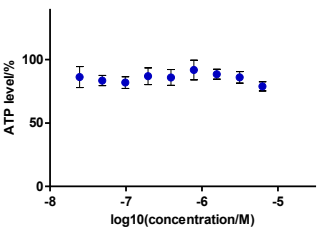
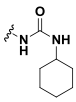
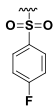
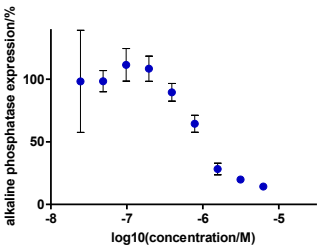
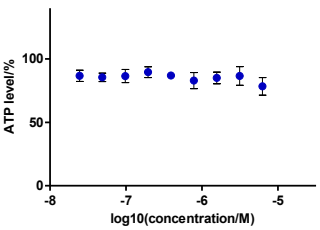
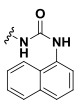
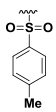
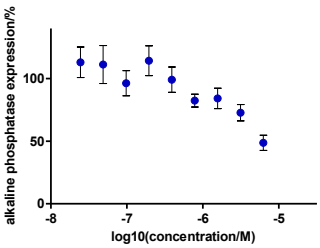
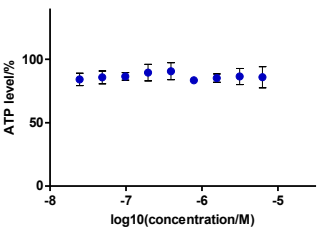
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
65	C						1.45	b	-
66	A						2.06	b	-
67	A						1.42	b	-
68	A						0.93	b	-
69	A						5.36	b	-

Table S3.1. (Continued)

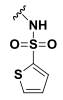
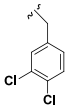
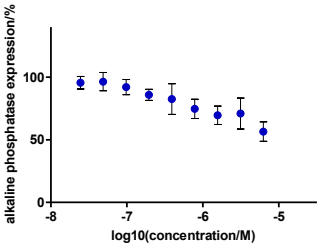
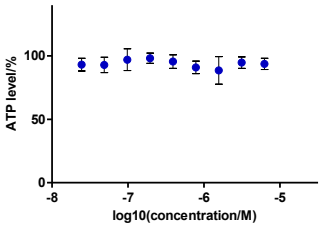
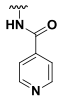
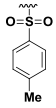
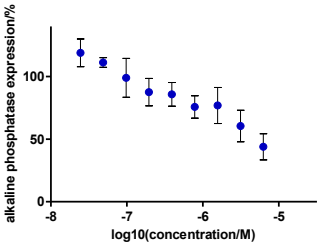
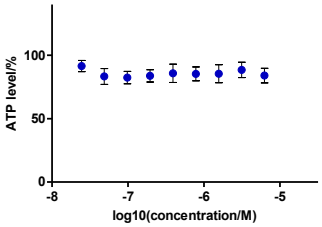
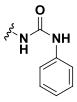
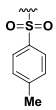
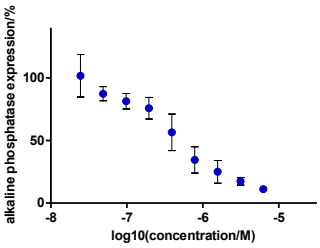
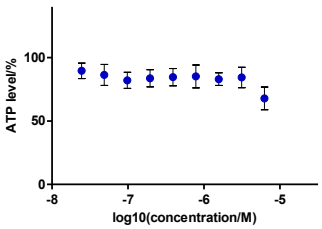
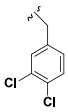
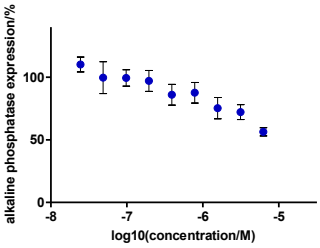
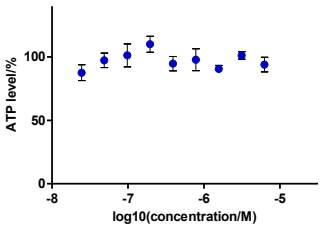
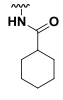
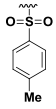
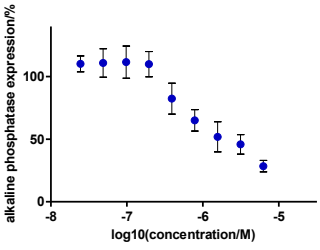
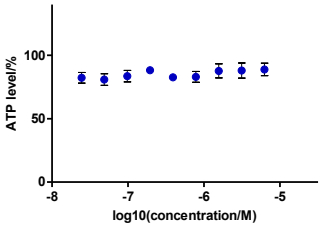
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
70	A						5.24	b	-
71	A						3.23	b	-
72	A						0.42	b	-
73	A						b	b	-
74	A						1.55	b	-

Table S3.1. (Continued)

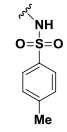
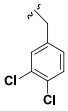
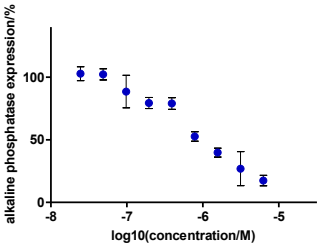
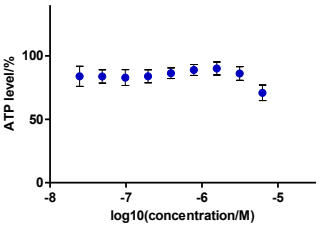
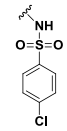
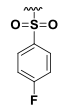
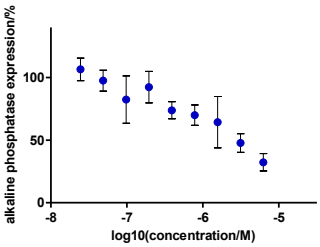
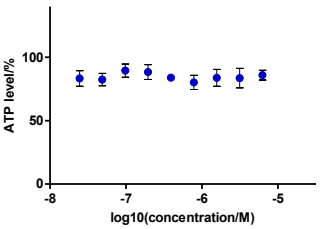
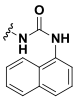
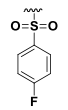
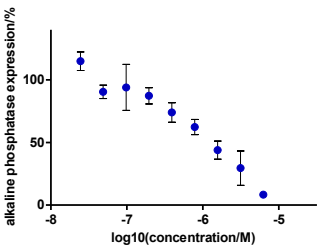
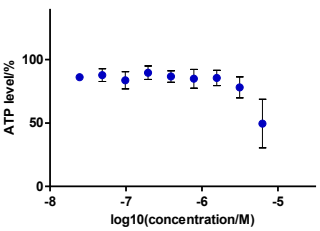
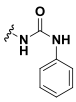
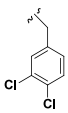
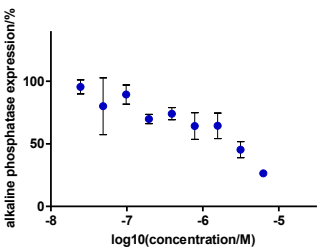
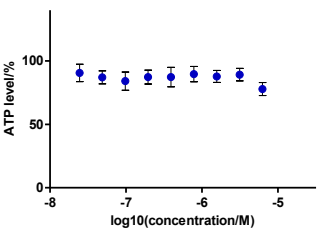
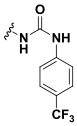
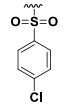
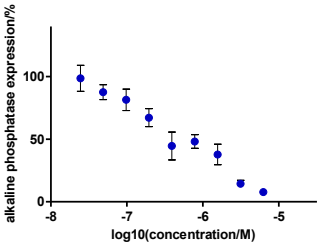
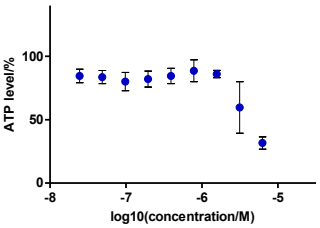
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
75	A						0.82	b	-
76	A						1.82	b	-
77	A						0.89	b	-
78	A						1.13	b	-
79	A						0.42	3.48	-

Table S3.1. (Continued)

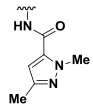
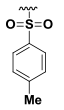
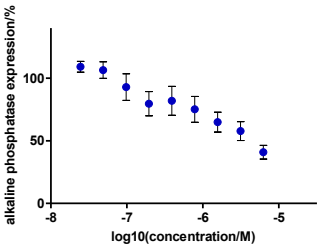
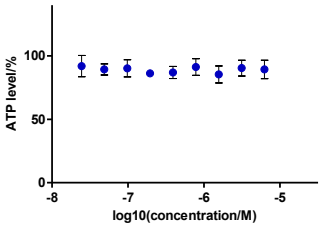
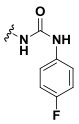
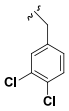
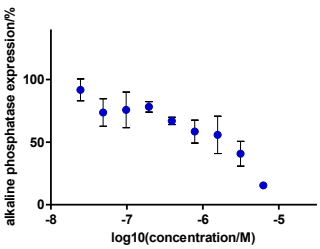
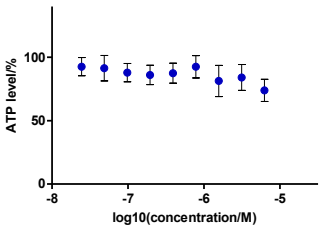
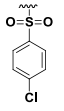
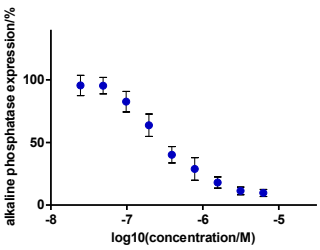
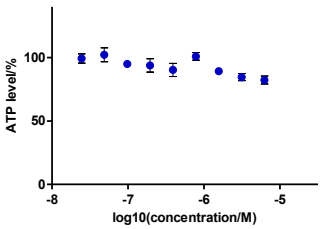
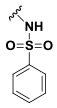
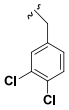
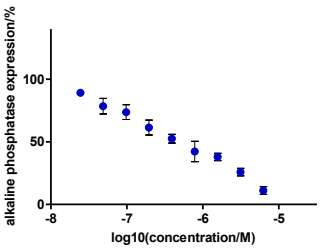
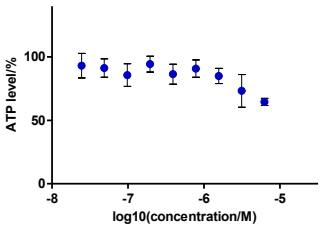
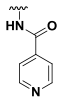
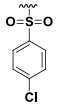
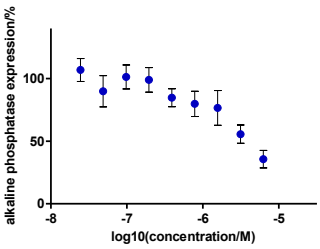
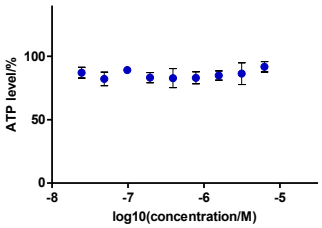
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
80	A						2.75	b	-
81	A						1.11	b	-
82	A						0.25	b	-
83	A						0.35	b	-
84	C						2.63	b	-

Table S3.1. (Continued)

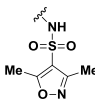
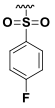
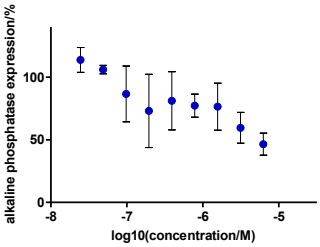
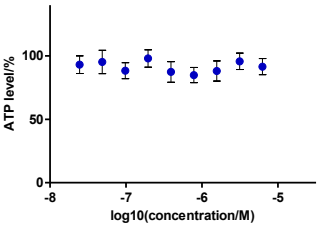
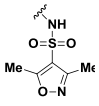
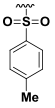
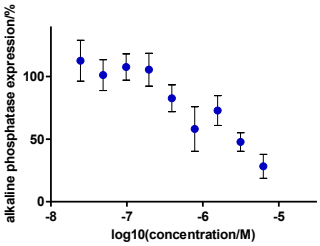
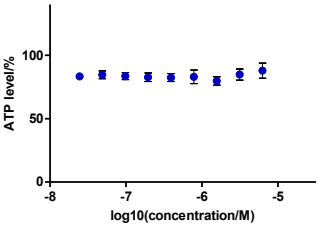
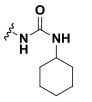
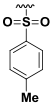
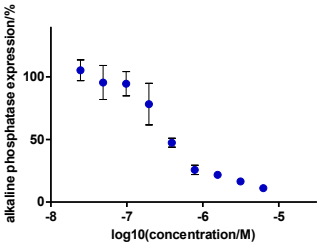
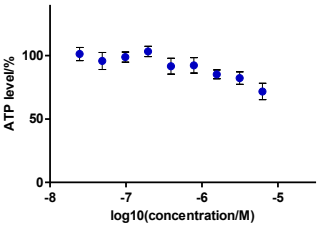
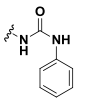
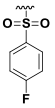
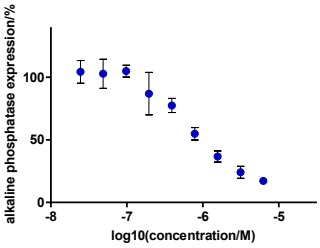
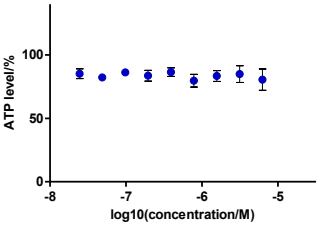
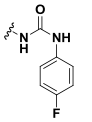
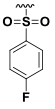
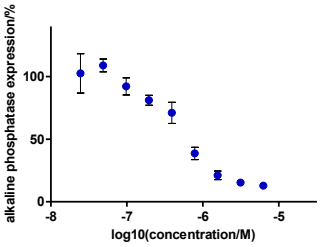
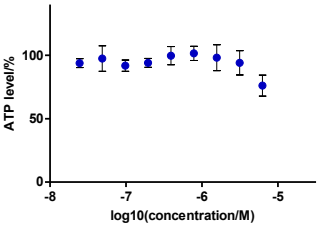
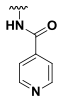
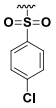
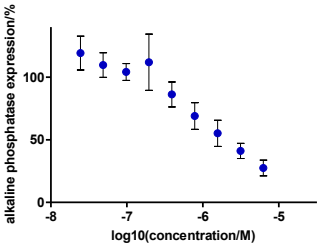
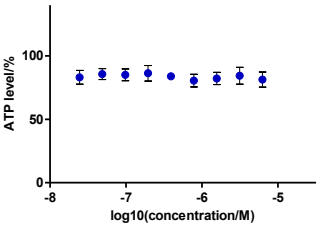
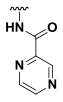
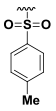
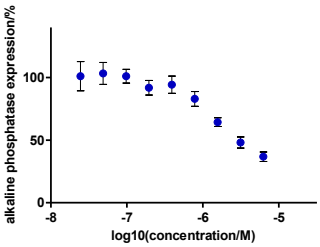
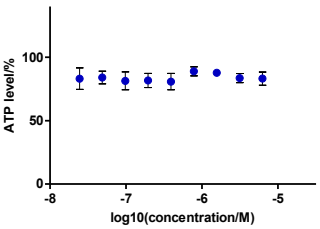
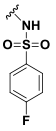
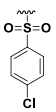
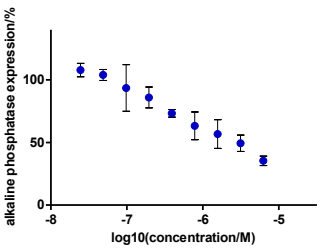
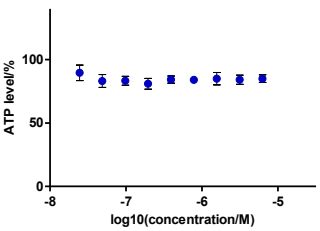
com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
85	A						4.52	b	-
86	A						2.02	b	-
87	A						0.30	b	-
88	A						0.86	b	-
89	A						0.49	b	-

Table S3.1. (Continued)

com-pound	scaf-fold #	R_x	R_y	R_z	osteoblast differentiation (CM activation)	viability in C3H10T1/2 cells (CM activation)	res-ponse EC_{50} (μM)	via-bilty EC_{50} (μM)	PBS solu-bility (μM)[a]
90	A						1.78	b	-
91	A						2.46	b	-
92	A						1.59	b	-

a) solubility was measured for compounds that were considered for subsequent experimentation based on their EC_{50}; b) dose-response curve did not pass EC_{50}, so an EC_{50} value was not calculated.